

# Sequential detection of multiple change points in networks: a graphical model approach

Arash Ali Amini and XuanLong Nguyen  
Department of Statistics, University of Michigan \*

July 9, 2012

## Abstract

We propose a probabilistic formulation that enables sequential detection of multiple change points in a network setting. We present a class of sequential detection rules for certain functionals of change points (minimum among a subset), and prove their asymptotic optimality properties in terms of expected detection delay time. Drawing from graphical model formalism, the sequential detection rules can be implemented by a computationally efficient message-passing protocol which may scale up linearly in network size and in waiting time. The effectiveness of our inference algorithm is demonstrated by simulations.

## 1 Introduction

Classical sequential detection is the problem of detecting changes in the distribution of data collected sequentially over time [2]. In a decentralized network setting, the decentralized sequential detection problem concerns with data sequences aggregated over the network, while sequential detection rules are constrained to the network structure (see, e.g., [3, 4, 5, 6, 7]). The focus was still on a *single* change point variable taking values in (discrete) time. In this paper, our interests lie in sequential detection in a network setting, where multiple change point variables may be simultaneously present.

As an example, quickest detection of traffic jams concerns with multiple potential hotspots (i.e., change points) spatially located across a highway network. A simplistic approach is to treat each change point variables independently, so that the sequential analysis of individual change points can be applied separately. However, it has been shown that accounting for the statistical dependence among the change point variables can provide significant improvement in reducing both false alarm probability and detection delay time [8].

This paper proposes a general probabilistic formulation for the multiple change point problem in a network setting, adopting the perspective of probabilistic graphical models for multivariate data [9]. We consider estimating functionals of multiple change points defined globally and locally across the network. The probabilistic formulation enables the borrowing of statistical strength from one network site (associated with a change point variable) to another. We

---

\*A preliminary version of this paper was presented at the International Symposium on Information Theory, Boston, Massachusetts, July 2012 [1]. We would like to thank Ram Rajagopal for valuable discussions and help during the course of this research. This work was supported in part by NSF grants CCF-1115769 and OCI-1047871.

propose a class of sequential detection rules, which can be implemented in a message-passing and distributed fashion across the network. The computation of the proposed sequential rules scales up linearly in both network size and in waiting time, while an approximate version scales up constantly in waiting time. The proposed detection rules are shown to be asymptotically optimal in a Bayesian setting. Interestingly, the expected detection delay can be expressed in terms of Kullback-Leibler divergences defined along edges of the network structure. We provide simulations that demonstrate both statistical and computational efficiency of our approach.

**Related Work.** The rich statistical literature on sequential analysis tends to focus almost entirely on the inference of a single change point variable [2, 10]. There are recent formulations for sequential diagnosis of a single change point, which may be associated with multiple causes [11], or multiple sequences [12]. Another approach taken in [13] considers a change propagating in a Markov fashion across an array of sensors. These are interesting directions but the focus is still on detecting the onset of a single event. Graphical models have been considered for distributed learning and decentralized detection before, but not in the sequential setting [14, 15]. This paper follows the line of work of [8, 16], but our formulation based on graphical models is more general, and we impose less severe constraints on the amount of information that can be exchanged across network sites.

**Notation.** We will use  $P$  to denote densities w.r.t. some underlying measure (usually understood from the context), while  $\mathbb{P}$  is used to denote probability measures.  $[d]$  denotes the set of integers  $\{1, \dots, d\}$ . For a real-valued function  $f$  defined on some space,  $\|f\|_\infty := \sup_x |f(x)|$  denotes its uniform norm. In an undirected graph, the neighborhood of a node  $i$  is denoted as  $\partial i$ .

## 2 Graphical model for multiple change points

In this section, we shall formulate the multiple change point detection problem, where the change point variables and observed data are linked using a graphical model. Consider a sensor network with  $d$  sensors, each of which is associated with a random variable  $\lambda_j \in \mathbb{N}$ , for  $j \in [d] := \{1, 2, \dots, d\}$ , representing a *change point*, the time at which a sensor fails to function properly. We are interested in detecting these change points as accurately and as early as possible, using the data that are associated with (e.g., observed by) the sensors. Taking a Bayesian approach, each  $\lambda_j$  is independently endowed with a prior distribution  $\pi_j(\cdot)$ .

A central ingredient in our formalism is the notion of a *statistical graph*, denoted as  $G = (V, E)$ , which specifies the probabilistic linkage between the change point variables and observed data collected in the network (cf. Fig. 1). The vertex set of the graph,  $V = [d]$  represents the indices of the change point variables  $\lambda_j$ . The edge set  $E$  represents pairings of change point variables,  $E = \{e = \{s_1, s_2\} \mid s_1, s_2 \in V\}$ . With each vertex and each edge, we associate a sequence of *observation* variables,

$$\mathbf{X}_j = (X_j^1, X_j^2, \dots), \quad j \in V, \quad (1)$$

$$\mathbf{X}_e = (X_e^1, X_e^2, \dots), \quad e \in E, \quad (2)$$

where the superscript denotes the time index. The  $\mathbf{X}_j$  models the private information of node  $j$ , while  $\mathbf{X}_e$  models the shared information of nodes connected by  $e$ . We will use the notation

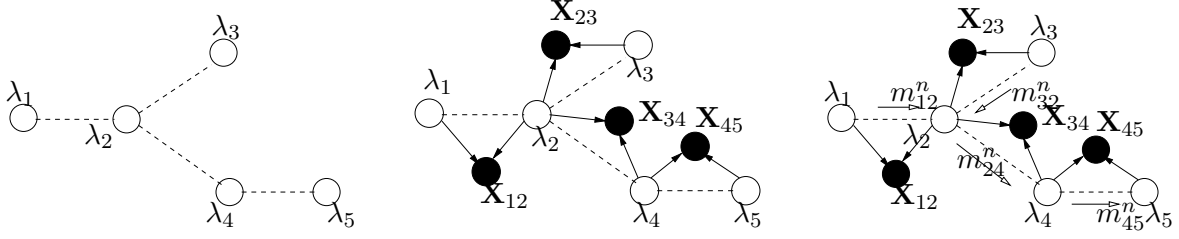


Figure 1: Left panel illustrates a statistical graph, which induces a graphical model in the middle panel. Right panel illustrates statistical messages passed at time  $n$  along some edges in a communication graph (which coincides with statistical graph in this case).

$\mathbf{X}_j^n = (X_j^1, \dots, X_j^n)$  and similarly for  $\mathbf{X}_e^n$ ; notice the distinction between  $X_j^n$ , the observation at time  $n$ , versus bold  $\mathbf{X}_j^n$ , the observations up to time  $n$ , both at node  $j$ . The aggregate of all the observations in the network is denoted as  $\mathbf{X}_* = (\mathbf{X}_j, j \in V, \mathbf{X}_e, e \in E)$ . Similarly,  $\mathbf{X}_*^n$  represents all the observations up to time  $n$ . We will also use  $\lambda_* = (\lambda_j, j \in V)$ .

The joint distribution of  $\lambda_*$  and  $\mathbf{X}_*^n$  is given by a graphical model,

$$P(\lambda_*, \mathbf{X}_*^n) = \prod_{j \in V} \pi_j(\lambda_j) \prod_{j \in V} P(\mathbf{X}_j^n | \lambda_j) \prod_{e \in E} P(\mathbf{X}_e^n | \lambda_{s_1}, \lambda_{s_2}). \quad (3)$$

Given  $\lambda_j = k$ , we assume  $X_j^1, \dots, X_j^{k-1}$  to be i.i.d. with density  $g_j$  and  $X_j^k, X_j^{k+1}, \dots$  to be i.i.d. with density  $f_j$ . Given  $(\lambda_{s_1}, \lambda_{s_2})$ , we assume that the distribution of  $\mathbf{X}_e^n$  only depends on  $\lambda_e := \lambda_{s_1} \wedge \lambda_{s_2}$ , the minimum of the two change points; hence we often write  $P(\mathbf{X}_e^n | \lambda_e)$  instead of  $P(\mathbf{X}_e^n | \lambda_{s_1}, \lambda_{s_2})$ . Given  $\lambda_e = k$ ,  $X_e^1, \dots, X_e^{k-1}$  are i.i.d. with density  $g_e$  and  $X_e^k, X_e^{k+1}, \dots$  are i.i.d. with density  $f_e$ . All the densities are assumed to be with respect to some underlying measure  $\mu$ . These specifications can be summarized as,

$$P(\mathbf{X}_j^n | \lambda_j) = \prod_{t=1}^{k-1} g_j(X_j^t) \prod_{t=k}^n f_j(X_j^t) \quad (4)$$

and similarly for  $P(\mathbf{X}_e^n | \lambda_e)$ . We will assume the prior on  $\lambda_j$  to be geometric with parameter  $\rho_j \in (0, 1)$ , i.e.  $\pi_j(k) := (1 - \rho_j)^{k-1} \rho_j$ , for  $k \in \mathbb{N}$ . Note that these change point variables are dependent a posteriori, despite being independent a priori.

## 2.1 Sequential rules and optimality

Although our primary interest is in sequential estimation of the change points  $\lambda_* = (\lambda_j)$ , we are in general interested in the following functionals,

$$\phi := \phi(\lambda_*) := \lambda_{\mathcal{S}} := \min_{j \in \mathcal{S}} \lambda_j. \quad (5)$$

for some subset  $\mathcal{S} \subset [d]$ . Examples include a single change point  $\mathcal{S} = \{j\}$ , the earliest among a pair  $\mathcal{S} = \{i, j\}$  and the earliest in the entire network  $\mathcal{S} = [d]$ . Let  $\mathcal{F}_n = \sigma(\mathbf{X}_*^n)$  be the  $\sigma$ -algebra induced by the sequence  $\mathbf{X}_*^n$ . A sequential detection rule for  $\phi$  is formally a stopping time  $\tau$  with respect to filtration  $(\mathcal{F}_n)_{n \geq 0}$ . To emphasize the subset  $\mathcal{S}$ , we will use  $\tau_{\mathcal{S}}$  to denote a rule when the functional  $\phi = \lambda_{\mathcal{S}}$ . For example  $\tau_1$  is a detection rule for  $\lambda_1$  and  $\tau_{12}$  is a rule for  $\lambda_{12} = \lambda_1 \wedge \lambda_2$ .

In choosing  $\tau$ , there is a trade-off between the false alarm probability  $\mathbb{P}(\tau \leq \phi)$  and the detection delay  $\mathbb{E}(\tau - \phi)_+$ . Here, we adopt the Neyman-Pearson setting to consider all stopping rules for  $\phi$ , having false alarm at most  $\alpha$ ,

$$\Delta_\phi(\alpha) := \{\tau : \mathbb{P}(\tau \leq \phi) \leq \alpha\}, \quad (6)$$

and pick a rule in  $\Delta_\phi$  that has minimum detection delay.

## 2.2 Communication graph and message passing (MP)

Another ingredient of our formalism is the notion of a *communication graph* representing constraints under which the data can be transmitted across network to compute a particular stopping rule, say  $\tau_j$ . In general, such a rule depends on all the aggregated data  $\mathbf{X}_*^n$ . We are primarily interested in those rules that can be implemented in a distributed fashion by passing messages from one sensor only to its neighbors in the communication graph. Although, conceptually, the statistical graph and communication graphs play two distinct roles, they usually coincide in practice and this will be assumed throughout this paper. See Fig. 1 for an illustration.

## 3 Proposed stopping rules

In general, we suspect that obtaining strictly optimal rules in closed form is not possible for the multiple change point problem introduced earlier; more crucially such rules are not computationally tractable for large networks. In this section, we shall present a class of detection rules that scale linearly in the size of the network,  $d$ , and can be implemented in a distributed fashion by message passing.

Consider the following posterior probabilities

$$\gamma_s^n(k) := \mathbb{P}(\lambda_s = k \mid \mathbf{X}_*^n), \quad (7)$$

$$\gamma_s^n[n] := \mathbb{P}(\lambda_s \leq n \mid \mathbf{X}_*^n) = \sum_{k=1}^n \gamma_s^n(k). \quad (8)$$

We propose to stop at the first time  $\gamma_s^n[n]$  goes above a threshold,

$$\tau_s = \inf\{n \in \mathbb{N} : \gamma_s^n[n] \geq 1 - \alpha\} \quad (9)$$

where  $\alpha$  is the maximum tolerable false alarm. It is easily verified that these rules have a false alarm at most  $\alpha$ .

**Lemma 1.** *For  $\phi = \lambda_s$ , the rule  $\tau_s \in \Delta_\phi(\alpha)$ .*

More interestingly, we will show in Section 4 that  $\tau_s$  is asymptotically optimal for detecting  $\lambda_s$ . First, let us look at two message-passing (MP) implementations of the stopping rule (9).

### 3.1 Exact message passing algorithm

It is relatively simple to adapt the well-established belief propagation algorithm, also known as sum-product, to the graphical model (3). The algorithm produces exact values of the posterior  $\gamma_s^n$ , as defined in (7), in the cases where  $G$  is a polytree (and provides a reasonable estimate otherwise.) In this section, we provide the details for  $\mathcal{S} = \{j\}$  or  $\mathcal{S} = \{i, j\} \in E$ .

One issue in adapting the algorithm is the possible infinite support of  $\gamma_s^n$ . Thanks to a “constancy” property of the likelihood, it is possible to lump all the states after  $n$  when computing  $\gamma_s^n[n]$ .

**Lemma 2.** *Let  $\{i_1, i_2, \dots, i_r\} \subset [d]$  be a distinct collection of indices. The function*

$$(k_1, k_2, \dots, k_r) \mapsto P(\mathbf{X}_*^n | \lambda_{i_1} = k_1, \lambda_{i_2} = k_2, \dots, \lambda_{i_r} = k_r)$$

*is constant over  $\{n+1, n+2, \dots\}^r$ .*

See Appendix A for the proof. The algorithm is invoked at each time step  $n$ , by passing messages between nodes according to the following protocol: a node sends a message to one of its neighbors (in  $G$ ) when and only when it has received messages from all its other neighbors. Message passing continues until any node can be linked to any other node by a chain of messages, assuming a connected graph. For a tree, this is usually achieved by designating a node as root and passing messages from the root to the leaves and then backwards.

The message that node  $j$  sends to its neighbor  $i$ , at time  $n$ , is denoted as  $m_{ji}^n = [m_{ji}^n(1), \dots, m_{ji}^n(n+1)] \in \mathbb{R}^{n+1}$  and computed as

$$m_{ji}^n(k) = \sum_{k'=1}^{n+1} \left\{ \tilde{\pi}_j^n(k') P(\mathbf{X}_j^n | k') P(\mathbf{X}_{ij}^n | k \wedge k') \prod_{r \in \partial j \setminus \{i\}} m_{rj}^n(k') \right\} \quad (10)$$

for  $k \in [n+1]$ , where

$$\tilde{\pi}_j^n(k) := \begin{cases} \pi_j(k) & \text{for } k \in [n] \\ \pi_j[n]^c = \sum_{k=n+1}^{\infty} \pi_j(k) & \text{for } k = n+1. \end{cases} \quad (11)$$

and  $\partial j$  is the neighborhood set of  $j$ . Once the message passing ends,  $\gamma_j^n$  and  $\gamma_{ij}^n$  are readily available. We have

$$\gamma_j^n(k) \propto \tilde{\pi}_j^n(k) P(\mathbf{X}_j^n | k) \prod_{r \in \partial j} m_{rj}^n(k), \quad k \in [n]. \quad (12)$$

It also holds for  $k = n+1$  if the LHS is interpreted as  $\gamma_j^n[n]^c$ .

The same messages can be used to compute  $\zeta_{ij}^n(k_1, k_2) := P(\lambda_i = k_1, \lambda_j = k_2 | \mathbf{X}_*^n)$  for  $\{i, j\} \in E$ . We have

$$\zeta_{ij}^n(k_1, k_2) \propto \psi_{ij}^n(k_1, k_2) \prod_{r \in \partial i \setminus \{j\}} m_{ri}^n(k_1) \prod_{r \in \partial j \setminus \{i\}} m_{rj}^n(k_2) \quad (13)$$

where

$$\psi_{ij}^n(k_1, k_2) := \tilde{\pi}_i^n(k_1) \tilde{\pi}_j^n(k_2) P(\mathbf{X}_i^n | k_1) P(\mathbf{X}_j^n | k_2) P(\mathbf{X}_{ij}^n | k_1 \wedge k_2) \quad (14)$$

for  $(k_1, k_2) \in [n]^2$ , from which  $\gamma_{ij}^n$  can be computed.

Let us summarize the steps of the message passing algorithm in the case of a tree:

---

Message passing algorithm to compute the posteriors  $\gamma_j^n[n]$  and  $\gamma_{ij}^n[n]$

---

At time each time  $n$ :

1. Designate a node of the tree, say node 1 as root and direct the edges to point away from root.
  2. Initialize messages  $m_{ji}^n \in \mathbb{R}^{n+1}$  (one for each directed edge  $j \rightarrow i$ ) to the all ones vector. Compute  $\tilde{\pi}_j^n(k)$  for  $k \in [n+1], j \in [d]$  according to (11).
  3. Pass messages  $m_{ji}^n$  from a node  $j$  to each of its descendants  $i$  (that is,  $i \in \partial j$  for which  $j \rightarrow i$  is a directed edge.) according to equation (10). Do this, recursively, starting from root ( $j = 1$ ) until you reach each of the leaves.
  4. Reverse the direction of the edges and repeat Step 3, this time starting from leaves and ending once you reached the root. In computing  $m_{ji}^n$  based on (10), use messages computed in the previous step.
  5. Compute  $\gamma_j^n(k)$  for  $k \in [n+1]$  based on (12) and normalize so that  $\sum_{k=1}^{n+1} \gamma_j^n(k) = 1$ . Let  $\gamma_j^n[n] = \sum_{k=1}^n \gamma_j^n(k)$ .
  6. Compute  $\zeta_{ij}^n(k_1, k_2)$  for  $(k_1, k_2) \in [n+1]^2$  based on (13) and (14) and normalize so that  $\sum_{k_1=1}^{n+1} \sum_{k_2=1}^{n+1} \zeta_{ij}^n(k_1, k_2) = 1$ . Let  $\gamma_{ij}^n[n] := \sum_{k_1=1}^n \sum_{k_2=1}^n \zeta_{ij}^n(k_1, k_2)$ .
- 

We have the following guarantee which is a restatement of a well-known result for belief propagation [17]:

**Lemma 3.** *When  $G$  is a tree, the message passing algorithm above produces correct values of  $\gamma_j^n$  and  $\gamma_{ij}^n$  at time step  $n$ , with computational complexity  $O((|V| + |E|)n)$ .*

## 4 Asymptotic optimality of MP rules

This section contains our main result on the asymptotic optimality of stopping rule (9). To simplify the statement of the results, let us extend the edge set to  $\tilde{E} := E \cup \{\{j\} : j \in V\}$ . This allows us to treat the private data associated with node  $j$ , i.e.  $\mathbf{X}_j$ , as (shared) data associated with a self-loop in the graph  $(V, \tilde{E})$ . For any  $e \in \tilde{E}$ , let  $I_e := \int f_e \log \frac{f_e}{g_e} d\mu$  be the KL divergence between  $f_e$  and  $g_e$ . For  $\phi = \lambda_S$ , let

$$I_\phi := I_{\lambda_S} := \sum_{e \in \mathcal{S}} I_e \quad (15)$$

where the sum runs over all  $e \in \tilde{E}$  which are subsets of  $\mathcal{S}$ . For example, for a chain graph on  $\{1, 2, 3\}$  with node 2 in the middle,  $\tilde{E} = \{\{1, 2\}, \{2, 3\}, \{1\}, \{2\}, \{3\}\}$  and we have  $I_{\lambda_{12}} := I_1 + I_2 + I_{12}$  while  $I_{\lambda_{13}} := I_1 + I_3$ . (Here, we abuse notation to write  $I_{12}$  instead of  $I_{\{1,2\}}$  and so on.)

Recall the geometric prior on  $\lambda_j$  (with parameter  $\rho_j$ ) and the definition of  $\phi = \lambda_S$  as the minimum of  $\lambda_j, j \in \mathcal{S}$ . Then,  $\phi$  is geometrically distributed a priori with parameter  $1 - e^{-q_\phi} := 1 - \prod_{j \in \mathcal{S}} (1 - \rho_j)$ .

We can now state our main result on asymptotic optimality.

**Theorem 1.** (*Optimal delay*) Assume  $\|\log \frac{f_e}{g_e}\|_\infty \leq M$  for all  $e \in \tilde{E}$ , and geometric priors for  $\{\lambda_j\}$ . Then,  $\tau_s$  is asymptotically optimal for  $\phi = \lambda_s$ ; more specifically, as  $\alpha \rightarrow 0$ ,

$$\begin{aligned}\mathbb{E}[\tau_s - \lambda_s \mid \tau_s \geq \lambda_s] &= \frac{|\log \alpha|}{q_{\lambda_s} + I_{\lambda_s}}(1 + o(1)) \\ &= \inf_{\tilde{\tau} \in \Delta_\phi(\alpha)} \mathbb{E}[\tilde{\tau} - \lambda_s \mid \tilde{\tau} \geq \lambda_s].\end{aligned}$$

*Remark 1.* Let us highlight some particular cases of interest in this result. To simplify notation, let  $\bar{\rho}_j := 1 - \rho_j$ .

- For  $\phi = \lambda_1 \wedge \dots \wedge \lambda_d$  (the minimum of all the change points), the asymptotic optimal delay is

$$\frac{|\log \alpha|}{-\sum_{j \in V} \log \bar{\rho}_j + \sum_{j \in V} I_j + \sum_{e \in E} I_e}(1 + o(1))$$

- For  $\phi = \lambda_i \wedge \lambda_j$ , the asymptotic optimal delay is

$$\frac{|\log \alpha|}{-\log \bar{\rho}_i - \log \bar{\rho}_j + I_i + I_j + I_{ij} 1_{\{\{i,j\} \in E\}}}(1 + o(1))$$

where  $1_{\{\{i,j\} \in E\}}$  is an indicator function, i.e., equal to 1 if  $\{i, j\}$  is an edge and zero otherwise.

- For  $\phi = \lambda_i$ , the asymptotic optimal delay is

$$\frac{|\log \alpha|}{-\log \bar{\rho}_i + I_i}(1 + o(1))$$

*Remark 2.* A particular feature of the asymptotic delay is the decomposition (15) of information along the edges of the graph. This is more clearly seen in the case of a paired delay  $\phi = \lambda_{ij}$ , for which the information  $I_\phi = I_i + I_j + I_{ij} 1_{\{\{i,j\} \in E\}}$  increases (hence the asymptotic delay decreases) if there is an edge between nodes  $i$  and  $j$ . This has no counterpart in the classical theory where one looks at change points independently.

*Remark 3.* Another feature of the result is observed for a single delay, say  $\phi = \lambda_1$ , where one has  $I_\phi = I_1$  regardless of whether there are edges between node 1 and the rest of the nodes. Thus, the asymptotic delay for the threshold rule which bases its decision on the posterior probability of  $\lambda_1$  given all the data in the network ( $\mathbf{X}_*^n$ ) is the same as the one which bases its decision on the posterior given only private data of node 1 ( $\mathbf{X}_1^n$ ). Although this rather counter-intuitive result holds asymptotically, the simulations show that even for moderately low values of  $\alpha$ , having access to extra information in  $\mathbf{X}_*^n$  does indeed improve performance as one expects. (cf. Section 5).

*Remark 4.* The assumptions of bounded likelihood ratios ( $\|\log \frac{f_e}{g_e}\|_\infty \leq M$ ) and geometric priors on  $\{\lambda_j\}$  are crucial for our proof technique. The geometric distribution can be relaxed to any distribution with exponential tails, but we cannot allow for more heavy-tailed priors. A brief explanation is provided after stating Theorem 2 in Section 6. This theorem is a key ingredient in our argument and relies heavily on these assumptions. Exponential tails assumption is also used in the decoupling Lemma 6.

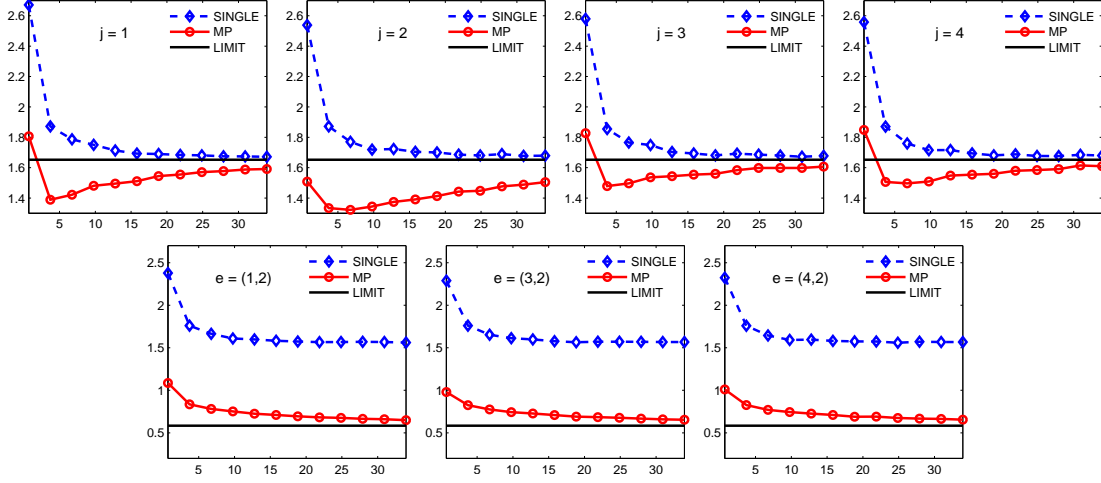


Figure 2: Plots of the slope  $\frac{1}{-\log \alpha} \mathbb{E}[\tau_S - \phi | \tau_S \geq \phi]$  against  $-\log \alpha$  for message-passing algorithm (MP) and SINGLE algorithm which disregards shared information. The graph is the star graph of 4 nodes with node 2 in the center. Estimates of both single and paired change points ( $\lambda_j$  and  $\lambda_{ij}$ ) are shown together with theoretical limit of Theorem 1. False alarm tolerance  $\alpha$  ranges in  $[0.5, 10^{-13}]$ .

## 5 Simulations

We present simulation results as depicted in Fig. 2. The setting is that of graphical model (3) on  $d = 4$  nodes, where the statistical graph is a star with node 2 in the middle. Conditioned on  $\lambda_*$ , all the data sequences,  $\mathbf{X}_*$ , are assumed Gaussian of variance 1, with pre-change mean 1 and post-change mean zero. All priors are geometric with parameters  $\rho_j = 0.1$ . Fig. 2 shows plots of expected delay over  $|\log \alpha|$ , against  $|\log \alpha|$ , for two methods: the message-passing algorithm of Section 3.1 (MP) and the method which bases its inference on posteriors calculated based only on each node's private information (SINGLE). This latter method estimates a single change point  $\lambda_j$  by  $\hat{\tau}_j := \inf\{n : P(\lambda_j \leq n | \mathbf{X}_j^n) \geq 1 - \alpha\}$  and a paired  $\lambda_{ij} = \lambda_i \wedge \lambda_j$  by  $\hat{\tau}_i \wedge \hat{\tau}_j$ . Also shown in the figure is the limiting value of the normalized expected delay as predicted by Theorem 1. All plots are generated by Monte Carlo simulation over 5000 realizations.

In estimating single change points, MP, which takes shared information into account, has a clear advantage over SINGLE, for high to relatively low false alarm values (even, say, around  $\alpha \approx e^{-5}$ ); though, both methods seem to converge to the same slope in the  $\alpha \rightarrow 0$  limit, as suggested by Theorem 1. (The particular value is  $(-\log 0.9 + 0.5)^{-1} = 1.6519$ .) Also note that the advantage of MP over SINGLE is more emphasized for node 2, as expected by its access to shared information from all the three nodes.

For paired change points, the advantage of MP over SINGLE is more emphasized. It is also interesting to note that while MP seems to converge to the expected theoretical limit  $(-2 \log 0.9 + 3 \cdot 0.5)^{-1} = 0.5845$ , SINGLE seems to converge to a higher slope (with a reasonable guess being 1.6519 as in the case of single change points).

In regard to false alarm probability, nonzero values were only observed for the first few values of  $\alpha$  considered here, and those were either below or very close to the specified tolerance.



## 6 Concentration inequalities for marginal likelihood ratios

In this section, we lay the groundwork for the proof of Theorem 1. The main result here is Theorem 2, which establishes concentration inequalities for various terms that appear in an asymptotic expansion of the marginal likelihood ratio, defined in (17) below. These terms (cf. (23) and (24)) are natural by-products of marginalization over a graph and their asymptotic behavior might be of independent interest.

Our standing assumption throughout is that the graph  $G = (V, E)$  is complete. This simplifies the arguments without loss of generality, since one can otherwise make the graph complete, by assigning sequences of i.i.d. data to each non-edge (with the same pre- and post-change distributions). These i.i.d. data do not affect the likelihood (as can be verified by examining the representation of Lemma 5) and they do not contribute to asymptotic delay since the corresponding KL informations are zero.

Fix some delay functional  $\phi = \tau_s$  throughout this section. We use the following notation regarding conditional probabilities and expectations

$$\begin{aligned}\mathbb{P}_\phi^k &:= \mathbb{P}(\cdot \mid \phi = k), & \mathbb{E}_\phi^k &:= \mathbb{E}(\cdot \mid \phi = k) \\ \mathbb{P}_{\lambda_*}^{m_*} &:= \mathbb{P}(\cdot \mid \lambda_* = m_*), & \mathbb{E}_{\lambda_*}^{m_*} &:= \mathbb{E}(\cdot \mid \lambda_* = m_*),\end{aligned}$$

for  $k \in \mathbb{N}$  and  $m_* = (m_1, \dots, m_d) \in \mathbb{N}^d$ . Here  $\{\lambda_* = m_*\} = \cap_{j=1}^d \{\lambda_j = m_j\}$ . Furthermore, let

$$\pi_\phi^k(m_*) := \mathbb{P}(\lambda_* = m_* \mid \phi = k). \quad (16)$$

Consider the marginal likelihood ratio

$$D_\phi^{k,n} := D_\phi^k(\mathbf{X}_*^n) := \frac{P(\mathbf{X}_*^n \mid \phi = k)}{P(\mathbf{X}_*^n \mid \phi = \infty)}. \quad (17)$$

Our asymptotic analysis hinges on the behavior of  $\frac{1}{n} \log D_\phi^k(\mathbf{X}_*^n)$  as  $n \rightarrow \infty$ , under probability measure  $\mathbb{P}_\phi^k$ . In particular, as a direct consequences of the results of [18], if one can show that

$$\mathbb{P}_\phi^k \left[ \frac{1}{N} \max_{1 \leq n \leq N} \log D_\phi^k(\mathbf{X}_*^{k+n}) \geq (1 + \varepsilon) I_\phi \right] \xrightarrow{N \rightarrow \infty} 0 \quad (18)$$

for all (small)  $\varepsilon > 0$  and all  $k \in \mathbb{N}$ , then the “lower bound” follows,  $\inf_{\tilde{\tau} \in \Delta_\phi(\alpha)} \mathbb{E}[\tilde{\tau} - \phi \mid \tilde{\tau} \geq \phi] \geq \frac{|\log \alpha|}{q_\phi + I_\phi} (1 + o(1))$ . Furthermore, let

$$T_\varepsilon^k := \sup \left\{ n \in \mathbb{N} : \frac{1}{n} \log D_\phi^k(\mathbf{X}_*^{k+n-1}) < I_\phi - \varepsilon \right\}.$$

By the results of [18], if one has

$$\mathbb{E} T_\varepsilon^\phi := \sum_{k=1}^{\infty} \mathbb{P}(\phi = k) \mathbb{E}_\phi^k(T_\varepsilon^k) < \infty, \quad (19)$$

for all (small)  $\varepsilon > 0$ , then the “upper bound” follows, that is,  $\tau_s$  as defined in (9) satisfies  $\mathbb{E}[\tau_s - \phi \mid \tau_s \geq \phi] \leq \frac{|\log \alpha|}{q_\phi + I_\phi} (1 + o(1))$ .

The following lemma provides sufficient conditions based on concentration inequalities under conditional probability measures  $\mathbb{P}_{\lambda_*}^{m_*}$ . In the following  $\varepsilon_0 > 0$  is some constant. (See Appendix D for the proof.)

**Lemma 4.** Assume that for all  $m_* \in \mathbb{N}^d$  for which  $\pi_\phi^k(m_*) > 0$ , one has

$$\mathbb{P}_{\lambda_*}^{m_*} \left\{ \left| \frac{1}{n} \log D_\phi^k(\mathbf{X}_*^n) - I_\phi \right| > \varepsilon \right\} \leq q(n) \exp(-c_1 n \varepsilon^2) \quad (20)$$

for all  $n \in \mathbb{N}$  and  $\varepsilon \in (0, \varepsilon_0)$  such that  $\sqrt{n} \geq \frac{1}{\varepsilon} p(m_*, k)$ , where  $p(\cdot)$  and  $q(\cdot)$  are polynomials with constant nonnegative coefficients. Furthermore, assume that both  $\pi_\phi^k(\cdot)$  and  $\mathbb{P}(\phi = \cdot)$  have finite polynomial moments. Then both (18) and (19) hold, hence Theorem 1 holds.

*Remark 1.* The condition of finite polynomial moments for  $\pi_\phi^k(\cdot)$  and  $\mathbb{P}(\phi = \cdot)$  is satisfied for a  $\phi = \min_{j \in \mathcal{S}} \lambda_j$  under geometric priors on  $\{\lambda_j\}$ .

In order to apply Lemma 4 easily, we introduce a notion of “stochastic asymptotic  $\varepsilon$ -equivalence” for sequence of random variables. To simplify notation, let  $\text{supp}(\pi_\phi^k) := \{m_* \in \mathbb{N}^d : \pi_\phi^k(m_*) > 0\}$ .

**Definition 1.** Consider two sequences  $\{a_n\}$  and  $\{b_n\}$  of random variables, where  $a_n = a_n(k)$  and  $b_n = b_n(k)$  could depend on a common parameter  $k \in \mathbb{N}$ . The two sequences are called “asymptotically  $\varepsilon$ -equivalent” as  $n \rightarrow \infty$ , w.r.t. the collection  $\{\mathbb{P}_{\lambda_*}^{m_*} : m_* \in \text{supp}(\pi_\phi^k)\}$ , and denoted

$$a_n \overset{\varepsilon}{\asymp} b_n,$$

if there exist polynomials  $p(\cdot)$  and  $q(\cdot)$  (with constant nonnegative coefficients), and  $\varepsilon_0 > 0$ , such that for all  $m_* \in \text{supp}(\pi_\phi^k)$ , we have

$$\mathbb{P}_{\lambda_*}^{m_*} (|a_n - b_n| \leq \varepsilon) \geq 1 - q(n) e^{-c_1 n \varepsilon^2}$$

for all  $n \in \mathbb{N}$  and  $\varepsilon \in (0, \varepsilon_0)$  satisfying  $\sqrt{n} \varepsilon \geq p(m_*, k)$ . The one-sided version, e.g.  $a_n \overset{\varepsilon}{\preceq} b_n$  is defined by replacing  $|a_n - b_n| \leq \varepsilon$  with  $a_n \leq b_n + \varepsilon$ . (The constants are independent of  $n, m_*, k$ , and  $\varepsilon$ , but they could depend on other parameters of the problem.)

By application of union bound and algebra, a finite number of asymptotic  $\varepsilon$ -equivalence statements can be manipulated under some algebraic rules to produce new such statements. Below, we summarize some of the rules:

(R1)  $a_n \overset{\varepsilon}{\asymp} b_n$  implies  $a_n \overset{C\varepsilon}{\asymp} b_n$  for  $C > 0$  and  $\alpha a_n \overset{\varepsilon}{\asymp} \alpha b_n$  for  $\alpha \in \mathbb{R}$ .

(R2)  $a_n \overset{\varepsilon}{\asymp} b_n$  and  $b_n \overset{\varepsilon}{\asymp} c_n$  implies  $a_n \overset{\varepsilon}{\asymp} c_n$ . (Transitivity)

(R3)  $a_n \overset{\varepsilon}{\asymp} b_n$  and  $c_n \overset{\varepsilon}{\asymp} d_n$  implies  $a_n \pm c_n \overset{\varepsilon}{\asymp} b_n \pm d_n$ .

(R4)  $a_n \overset{\varepsilon}{\asymp} b_n$  implies  $\max\{a_n, c_n\} \overset{\varepsilon}{\asymp} \max\{b_n, c_n\}$ .

(R5)  $a_n \overset{\varepsilon}{\asymp} b_n$ ,  $c_n \overset{\varepsilon}{\asymp} 1$  and  $\{b_n\}$  bounded implies  $a_n c_n \overset{\varepsilon}{\asymp} b_n$ .

(R6)  $a_n \overset{\varepsilon}{\asymp} a > 0$  and  $b_n \overset{\varepsilon}{\preceq} -b < 0$  implies  $\max\{a_n, b_n\} \overset{\varepsilon}{\asymp} a$ .

(R7) “log-sum-max” inequality for positive sequences  $\{a_n\}$  and  $\{b_n\}$ :

$$n^{-1} \log(a_n + b_n) \overset{\varepsilon}{\asymp} \max\{n^{-1} \log a_n, n^{-1} \log b_n\}. \quad (21)$$

The last statement follows from inequalities  $0 \leq \log(a_n + b_n) - \max\{\log a_n, \log b_n\} \leq \log 2$ . Dividing by  $n$ , we observe that the difference is bounded by  $\varepsilon$ , in absolute value, as long as  $n\varepsilon \geq \log 2$ . This implies the condition in Definition 1, since  $\{(n, \varepsilon) : \sqrt{n}\varepsilon \geq \log 2\} \subset \{(n, \varepsilon) : n\varepsilon \geq \log 2\}$ .

As another example of how these rules are obtained, consider (R3). We have  $|a_n - b_n| \leq \varepsilon$  on event  $A_{1,n}$  having probability at least  $1 - q_1(n)e^{-c_1 n \varepsilon^2}$ , for  $\sqrt{n}\varepsilon \geq p_1(m_*, k)$ . Similarly,  $|b_n - c_n| \leq \varepsilon$  on event  $A_{2,n}$  with probability at least  $1 - q_2(n)e^{-c_2 n \varepsilon^2}$ , for  $\sqrt{n}\varepsilon \geq p_2(m_*, k)$ . Then, by union bound  $A_{1,n} \cap A_{2,n}$  has probability at least  $1 - (q_1(n) + q_2(n))e^{-(c_1 \wedge c_2)n\varepsilon^2}$ , for  $\sqrt{n}\varepsilon \geq p_1(m_*, k) + p_2(m_*, k)$ . For this range of  $n$ , on event  $A_{1,n} \cap A_{2,n}$ , we have both  $|a_n - b_n| \leq \varepsilon$  and  $|b_n - c_n| \leq \varepsilon$ , from which it follows  $|a_n - c_n| \leq 2\varepsilon$ , by triangle inequality. Since both  $q_1 + q_2$  and  $p_1 + p_2$  are polynomials, we have the desired assertion.

*Remark 1* According to Definition 1 and Lemma 4, to prove Theorem 1, it is enough to show that

$$\frac{1}{n} \log D_\phi^{k,n} \stackrel{\varepsilon}{\asymp} I_\phi \quad \text{as } n \rightarrow \infty \text{ w.r.t. } \{\mathbb{P}_{\lambda_*}^{m_*}\}$$

(We often omit  $m_* \in \text{supp}(\pi_\phi^k)$  when it is implicitly understood.) The rules stated above allows one to reduce the problem to asymptotic  $\varepsilon$ -equivalence statements for simpler terms, as considered in the next section. In this context, we regard parameters of the priors,  $\{\rho_j\}$ , and pre- and post-change densities as constants. In other words, the constants in the definition of  $\varepsilon$ -equivalence can depend on  $\{\rho_j\}$ ,  $\{I_e\}$ , and  $M$  (the uniform norm of  $\log(f_e/g_e)$ ).

We now introduce a couple of building blocks occurring frequently and establish  $\stackrel{\varepsilon}{\asymp}$  statements for them. Recall that  $f_e$  and  $g_e$  denote the pre- and post-change densities for edge  $e \in \tilde{E}$ . Define

$$R_k^n(e) := R_k^n(\mathbf{X}_e) := \prod_{t=k}^n \frac{f_e}{g_e}(\mathbf{X}_e) = \prod_{t=k}^n e^{h_e(\mathbf{X}_e)}, \quad h_e := \log \frac{f_e}{g_e}. \quad (22)$$

Note that by assumption  $\|h_e\|_\infty \leq M$  for all  $e$ . We will use the convention that empty products evaluate to 1, that is,  $R_k^n(e) = 1$  whenever  $k > n$ . We also define  $S$ -terms as

$$S_u^{\nu,n}(e) := \sum_{p=u}^{\nu} A e^{-\beta p} R_p^n(e) \quad (23)$$

where  $A$  and  $\beta$  are some positive constants. Similarly, define  $M$  and  $L$ -terms as follows

$$M_u^{\nu,n}(e) := \sum_{p_1=u}^{\nu} \sum_{p_2=u}^{\nu} A e^{-(\beta_1 p_1 + \beta_2 p_2)} R_{p_1 \wedge p_2}^n(e) \quad (24)$$

$$L_{u,(r)}^{\nu,n}(e) := \sum_{p=u}^{\nu} A e^{-\beta p} R_{p \wedge r}^n(e) \quad (25)$$

for constants  $A, \beta_1, \beta_2, \beta > 0$ . The constants involved in these definitions can be different in each occurrence and we have suppressed them in the notation for simplicity. The  $M$  and  $L$ -terms are most relevant when  $e$  is a proper edge, that is,  $e = \{i, j\} \in \tilde{E}$  and  $i \neq j$ , although the statements involving them hold in general.

The following lemma is proved in Section 8. Recall that  $I_e$  is the KL divergence between  $f_e$  and  $g_e$ , that is,  $I_e := \int f_e \log \frac{f_e}{g_e}$ .

**Theorem 2.** Assume  $\|\log \frac{f_e}{g_e}\|_\infty \leq M$  for all  $e \in \tilde{E}$ . The following asymptotic  $\varepsilon$ -equivalence relations hold with respect to  $\{\mathbb{P}_{\lambda_*}^{m_*} : m_* \in \text{supp}(\pi_\phi^k)\}$ , as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \log R_u^n(e) \stackrel{\varepsilon}{\asymp} \frac{1}{n} \log S_u^{\infty,n}(e) \stackrel{\varepsilon}{\asymp} \frac{1}{n} \log M_u^{\infty,n}(e) \stackrel{\varepsilon}{\asymp} \frac{1}{n} \log L_{u,(r)}^{\infty,n}(e) \stackrel{\varepsilon}{\asymp} I_e \quad (26)$$

for any  $u, r \leq 2k$  and  $e \in \tilde{E}$ .

The proof of this theorem is deferred to Section 8. The log  $\varepsilon$ -equivalence  $\frac{1}{n} \log R_u^n(e) \stackrel{\varepsilon}{\asymp} I_e$  is intuitive as will become clear in the proof. The lemma essentially states that there are no surprises regarding  $S$ ,  $M$  and  $L$  terms and they are all  $\varepsilon$ -equivalent to the corresponding edge information. We also note that  $2k$  in the statement of the Lemma can be replaced with  $Ck$  for any constant  $C > 0$ .

*Remark.* Let us consider the role of our assumptions on the priors and likelihood ratios, by giving a high-level overview of the proof of Theorem 2 for  $S_u^{\infty,n}(e)$ . The exponential decay for the tails of the priors is reflected in the definition of  $S_u^{\infty,n}(e)$  in (23). The terms  $R_p^n(e)$  in this sum are concentrated around  $I_e$  if  $p \ll n$  (as in this case  $R_p^n(e)$  is the product of many essentially i.i.d. terms). For  $p$  close to  $n$ , however, there is no guaranteed concentration for  $R_p^n(e)$ , as it is a product of only a few random variables. For these terms, however, the prefactor  $e^{-\beta p}$  is small while  $R_p^n(e)$  is guaranteed to be bounded (based on  $\|\log(f_e/g_e)\|_\infty \leq M$ ). Hence these terms are negligible and do not contribute to  $\frac{1}{n} \log S_u^{\infty,n}(e)$ , asymptotically. This argument is made precise in Section 8.

To simplify notation, from now on, we will drop the second upper index in the symbols for  $S$ ,  $L$  and  $M$  terms, whenever this index is  $n$  and there is no chance of confusion. That is, we adhere to the following convention,

$$S_u^\nu(e) := S_u^{\nu,n}(e), \quad L_{u,(r)}^\nu(e) := L_{u,(r)}^{\nu,n}(e), \quad M_u^\nu(e) := M_u^{\nu,n}(e). \quad (27)$$

## 7 Proof of the optimal delay theorem

Let us define

$$\mathcal{M}_\phi^{k,n} := \mathcal{M}_\phi^k(\mathbf{X}_*^n) := \sum_{k_1, \dots, k_d} \pi_\phi^k(k_1, \dots, k_d) \prod_{j \in V} R_{k_j}^n\{j\} \prod_{e \in E} R_{k_e}^n(e) \quad (28)$$

where  $k_e := k_i \wedge k_j$  for  $e = \{i, j\}$ , and each variable  $k_j$  runs over  $\{1, 2, \dots\} \cup \{\infty\}$ . The inclusion of  $\infty$  in range of the summations does not affect the case  $k < \infty$ , but will allow us to use the same expression (28) for  $\mathcal{M}_\phi^{\infty,n}$ . We have following easily verified representation of  $D_\phi^{k,n}$ . (See Appendix B for the proof).

**Lemma 5.** With  $D_\phi^{k,n}$  defined as in (17),

$$D_\phi^{k,n} = \frac{\mathcal{M}_\phi^{k,n}}{\mathcal{M}_\phi^{\infty,n}}. \quad (29)$$

We will use the following technical lemma to decouple sums of products. Let  $\binom{[r]}{2}$  denote the collection of 2-subsets of  $[r] = \{1, \dots, r\}$ , with the convention that each member is a denoted as an ordered pair  $(i, j)$  with  $i < j$ .

**Lemma 6.** *Let  $\mathbf{S} = S_1 \times S_2 \times \dots \times S_r$  be the Cartesian product of  $r$  countable sets  $S_1, \dots, S_r$  and let  $\mathbf{k} = (k_1, \dots, k_r)$  be a multi-index for  $\mathbf{S}$ . Let  $F_j$  and  $G_{ij}$  be nonnegative functions defined on  $S_j$  and  $S_i \times S_j$  respectively, for  $i, j \in [r]$ . Let  $H_1$  be a nonnegative function on  $S_1$ . Let  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r) \in \mathbb{R}_{++}^r$ . Then,*

$$(a) \quad \sum_{k_1 \in S_1} e^{-\beta_1 k_1} F_1(k_1) H_1(k_1) \leq \left( \sum_{k_1 \in S_1} e^{-\beta_1 k_1/2} F_1(k_1) \right) \left( \sum_{k_1 \in S_1} e^{-\beta_1 k_1/2} H_1(k_1) \right). \quad (30)$$

$$(b) \quad \sum_{\mathbf{k} \in \mathbf{S}} \left\{ e^{-\boldsymbol{\beta}^T \mathbf{k}} \prod_{j=1}^r F_j(k_j) \prod_{(i,j) \in \binom{[r]}{2}} G_{ij}(k_i, k_j) \right\} \leq \left( \prod_{j=1}^r \left\{ \sum_{k_j \in S_j} e^{-\beta_j k_j/r} F_j(k_j) \right\} \right) \times \prod_{(i,j) \in \binom{[r]}{2}} \left\{ \sum_{(k_i, k_j) \in S_i \times S_j} e^{-(\beta_i k_i + \beta_j k_j)/r} G_{ij}(k_i, k_j) \right\} \quad (31)$$

The key in this lemma is that the functions  $F_j$ ,  $G_{ij}$  and  $H_1$  are nonnegative. One might already see how the application of Lemma 6 to the sum in (28) produces  $S$  and  $M$  terms as introduced in Section 6. We are ready to give the proof of Theorem 1. We start with the two extreme change point functionals  $\lambda_s$ : a single change point ( $|\mathcal{S}| = 1$ ), and the minimum of all the change points ( $\mathcal{S} = [d]$ ). Then, we present the proof for  $\lambda_s$  with  $1 < |\mathcal{S}| < d$ , omitting some of the details for brevity.

### 7.1 Proof for the case $\phi = \lambda_1 \wedge \dots \wedge \lambda_d$

First, note that in this case  $\mathcal{M}_\phi^{\infty, n} = 1$ , since  $\phi = \infty$  implies  $\lambda_1 = \dots = \lambda_d = \infty$ . Hence, we only need to consider  $\mathcal{M}_\phi^{k, n}$  for some  $k < \infty$ . We then observe that  $\pi_\phi^k(k_1, \dots, k_d)$  is nonzero only when at least one of  $k_1, \dots, k_d$  is equal to  $k$ . We break up the sum according to how many of  $k_1, \dots, k_d$  are equal to  $k$ .

Let  $\mathcal{J}$  be a subset of  $[d]$  of size  $|\mathcal{J}| = s$ . Let  $\mathcal{J}^c = [d] \setminus \mathcal{J}$ . Consider the terms in the sum (28) for which  $k_j = k$  for  $j \in \mathcal{J}$  and  $k_j > k$  for  $j \in \mathcal{J}^c$ . We call the sum over these terms  $T_{\mathcal{J}}$ . Then,  $\mathcal{M}_\phi^{k, n} = \sum_{\mathcal{J}: |\mathcal{J}| \geq 1} T_{\mathcal{J}} = \sum_{s=1}^d \sum_{\mathcal{J}: |\mathcal{J}|=s} T_{\mathcal{J}}$ , where the sum is over all subsets  $\mathcal{J}$  of  $[d]$  of size at least 1.

Let us fixed some  $s \in [d]$  and some  $\mathcal{J} \subset [d]$  with  $|\mathcal{J}| = s$ . Without loss of generality, we can pick  $\mathcal{J} = \{1, \dots, s\}$ . We note that

$$\begin{aligned} \pi_\phi^k(k, \dots, k, k_{s+1}, \dots, k_d) &= A \prod_{j=s+1}^d \bar{\rho}_j^{k_j} \\ &= A e^{-\sum_{j \in \mathcal{J}^c} \beta_j k_j}, \quad \text{for } k_j > k, j \in \mathcal{J}^c \end{aligned}$$

where  $\beta_j = -\log \bar{\rho}_j > 0$ , and  $A = A(\{\rho_j\})$  is some constant. It follows that

$$T_{\mathcal{J}} = \prod_{j \in \mathcal{J}} R_{k_e}^n\{j\} \prod_{|e \cap \mathcal{J}| \geq 1} R_k^n(e) \underbrace{\sum_{k_j > k, j \in \mathcal{J}^c} \left\{ A e^{-\sum_{j \in \mathcal{J}^c} \beta_j k_j} \prod_{j \in \mathcal{J}^c} R_{k_j}^n\{j\} \prod_{|e \cap \mathcal{J}|=0} R_{k_e}^n(e) \right\}}_{(*)}. \quad (32)$$

Here and in the rest of the proof, the index  $e$  runs in the set  $E$  of original edges (not the modified set  $\tilde{E}$  introduced in Section 4). That is, each edge  $e = \{i, j\} \in E$  for some  $i \neq j$ . Note that in (32), the rightmost product is over all 2-subset of  $\mathcal{I}^c$ , which we denote as  $\binom{\mathcal{I}^c}{2}$ . We can now apply first part of Lemma 6, with  $r = |\mathcal{I}^c| = d - s$ , to obtain

$$(\star) \leq A \prod_{j \in \mathcal{I}^c} \underbrace{\left\{ \sum_{k_j > k} e^{-\frac{\beta_j k_j}{d-s}} R_{k_j}^n \{j\} \right\}}_{(\star\star)} \prod_{(i,j) \in \binom{\mathcal{I}^c}{2}} \underbrace{\left\{ \sum_{\substack{k_i > k, \\ k_j > k}} e^{-\frac{\beta_i k_i + \beta_j k_j}{d-s}} R_{k_i \wedge k_j}^n \{i, j\} \right\}}_{(\star\star\star)}.$$

Each term denoted as  $(\star\star)$  is of the form  $S_{k+1}^\infty \{j\}$  and each term denoted as  $(\star\star\star)$  is of the form  $M_{k+1}^\infty \{i, j\}$ . Hence, we have

$$T_{\mathcal{I}} \leq A \prod_{j \in \mathcal{I}} R_k^n \{j\} \prod_{|e \cap \mathcal{I}| \geq 1} R_k^n(e) \prod_{j \in \mathcal{I}^c} S_{k+1}^\infty \{j\} \prod_{|e \cap \mathcal{I}|=0} M_{k+1}^\infty(e).$$

Applying Theorem 2 to each of the  $R, S$  and  $M$  forms above, we obtain

$$\begin{aligned} \frac{1}{n} \log T_{\mathcal{I}} &\stackrel{\varepsilon}{\preceq} \sum_{j \in \mathcal{I}} I_j + \sum_{|e \cap \mathcal{I}| \geq 1} I_e + \sum_{j \in \mathcal{I}^c} I_j + \sum_{|e \cap \mathcal{I}|=0} I_e \\ &= \sum_{j \in V} I_j + \sum_{e \in E} I_e. \end{aligned} \tag{33}$$

where the  $\varepsilon$ -equivalence in the above and in what follows is w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .

To obtain the lower bound, we bound  $(\star)$  from below by its first term,

$$(\star) \geq A e^{-(k+1) \sum_{j \in \mathcal{I}^c} \beta_j} \prod_{j \in \mathcal{I}^c} R_{k+1}^n \{j\} \prod_{|e \cap \mathcal{I}|=0} R_{k+1}^n(e)$$

which, after applying Theorem 2, gives us a lower bound on  $\frac{1}{n} \log T_{\mathcal{I}}$  matching the RHS of (33). Finally, note that the RHS of (33) does not depend on the particular choice of  $\mathcal{I}$ . We now use the log-sum-max rule (R7) to get

$$\begin{aligned} \frac{1}{n} \log \mathcal{M}_\phi^{k,n} &= \frac{1}{n} \log \left( \sum_{\mathcal{I}: |\mathcal{I}| \geq 1} T_{\mathcal{I}} \right) \\ &\stackrel{\varepsilon}{\succeq} \max_{\mathcal{I}: |\mathcal{I}| \geq 1} \left\{ \frac{1}{n} \log T_{\mathcal{I}} \right\} \stackrel{\varepsilon}{\succeq} \sum_{j \in V} I_j + \sum_{e \in E} I_e, \end{aligned}$$

which is the desired result.

## 7.2 Proof for the case $\phi = \lambda_1$

In this case, one has  $\pi_\phi^k(k_1, \dots, k_d) = 1\{k_1 = k\} \prod_{j=2}^d \pi_j(k_j)$ , hence

$$\pi_\phi^k(k_1, \dots, k_d) = \left[ 1\{k_1 = k\} A \prod_{j=2}^d \bar{\rho}_j^{k_j} \right] = \left[ 1\{k_1 = k\} A e^{-\sum_{j=2}^d \beta_j k_j} \right]$$

where  $\beta_j = -\log \bar{\rho}_j > 0$  and  $A = A(\{\rho_j\}) > 0$  is some constant. Then, we can write

$$\mathcal{M}_\phi^{k,n} = R_k^n\{1\} \sum_{k_2, \dots, k_d} \left\{ A e^{-\sum_{j=2}^d \beta_j k_j} \prod_{j=2}^d \left( R_{k_j}^n\{j\} R_{k \wedge k_j}^n\{1, j\} \right) \prod_{e: 1 \notin e} R_{k_e}^n(e) \right\}. \quad (34)$$

Note that the second product runs over all 2-subsets of  $[2 : d] := \{2, \dots, d\}$  which we denote as  $\binom{[2:d]}{2}$ . Hence, we can apply Lemma 6 with  $r = d - 1$  to obtain

$$\frac{\mathcal{M}_\phi^{k,n}}{AR_k^n\{1\}} \leq \prod_{j=2}^d \left\{ \sum_{k_j} e^{-\beta_j k_j / (d-1)} R_{k_j}^n\{j\} R_{k \wedge k_j}^n\{1, j\} \right\} \prod_{(i,j) \in \binom{[2:d]}{2}} \left\{ \sum_{k_i, k_j} e^{-(\beta_i k_i + \beta_j k_j) / (d-1)} R_{k_i \wedge k_j}^n\{i, j\} \right\}.$$

Each term appearing in second product is of the form  $M_1^\infty\{i, j\}$ . Applying the second half of Lemma 6 to the first product, we get

$$\frac{\mathcal{M}_\phi^{k,n}}{AR_k^n\{1\}} \leq \prod_{j=2}^d \underbrace{\left\{ \left( \sum_{k_j} e^{-\frac{\beta_j k_j}{2(d-1)}} R_{k_j}^n\{j\} \right) \right\}}_{(*)} \underbrace{\left\{ \left( \sum_{k_j} e^{-\frac{\beta_j k_j}{2(d-1)}} R_{k \wedge k_j}^n\{1, j\} \right) \right\}}_{(**)} \prod_{(i,j) \in \binom{[2:d]}{2}} M_1^\infty\{i, j\} \quad (35)$$

Each term denoted as  $(*)$  is of the form  $S_1^\infty\{j\}$ . For  $k < \infty$ , each term denoted as  $(**)$  can be written in the form  $L_{1,(k)}^\infty\{1, j\}$ . That is,

$$\frac{\mathcal{M}_\phi^{k,n}}{AR_k^n\{1\}} \leq \prod_{j=2}^d \left\{ S_1^\infty\{j\} L_{1,(k)}^\infty\{1, j\} \right\} \prod_{(i,j) \in \binom{[2:d]}{2}} M_1^\infty\{i, j\}$$

Applying Theorem 2 to each of the  $R$ ,  $S$ ,  $L$  and  $M$  forms above, we obtain

$$\frac{1}{n} \log \mathcal{M}_\phi^{k,n} \stackrel{\varepsilon}{\preceq} I_1 + \sum_{j=2}^d (I_j + I_{1,j}) + \sum_{(i,j) \in \binom{[2:d]}{2}} I_{ij} \quad (36)$$

where the  $\varepsilon$ -equivalence in the above and in what follows is w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m*}\}$ .

The lower bound is obtained, as in Section 7.1, by bounding the sum in (34) by its first term (i.e.,  $k_1 = k_2 = \dots = k_d = 1$ )

$$\mathcal{M}_\phi^{k,n} \geq R_k^n\{1\} A e^{-\sum_{j=2}^d \beta_j} \prod_{j=2}^d R_1^n\{j\} R_1^n\{1, j\} \prod_{(i,j) \in \binom{[2:d]}{2}} R_1^n\{i, j\}.$$

Applying  $\frac{1}{n} \log(\cdot)$  and using Theorem 2 for each term, we get a lower bound matching the RHS of (36). That is, the bound in (36) holds with  $\stackrel{\varepsilon}{\preceq}$  replaced with  $\stackrel{\varepsilon}{\simeq}$ .

Now consider the denominator of  $D_\phi^{k,n}$ , namely  $\mathcal{M}_\phi^{\infty,n}$ . An upper bound on  $\mathcal{M}_\phi^{\infty,n}$  can be obtained by letting  $k = \infty$  in (35). We note that  $R_\infty^n\{1\} = 1$  and that  $(**)$  is now a term of the form  $S_1^\infty\{1, j\}$ . Proceeding as before, we obtain an upper bound similar to that of (36), with  $I_1$  missing from the bound. The lower bound is obtained by the same technique. Hence,

$$\frac{1}{n} \log \mathcal{M}_\phi^{\infty,n} \stackrel{\varepsilon}{\simeq} \sum_{j=2}^d (I_j + I_{1,j}) + \sum_{(i,j) \in \binom{[2:d]}{2}} I_{ij}. \quad (37)$$

Combining equality form of (36) and (37), we have

$$\frac{1}{n} \log D_\phi^{k,n} = \frac{1}{n} \log \mathcal{M}_\phi^{k,n} - \frac{1}{n} \log \mathcal{M}_\phi^{\infty,n} \stackrel{\varepsilon}{\asymp} I_1 \quad (38)$$

which is the desired result.

### 7.3 Proof for $\lambda_S$ with $1 < |S| < d$

We now briefly give the proof for the remaining cases. Without loss of generality, we assume  $S = \{1, 2, \dots, r\}$  for some  $r \in \{2, \dots, d-1\}$ . In other words, the delay functional is  $\phi = \lambda_S = \lambda_1 \wedge \lambda_2 \wedge \dots \wedge \lambda_r$ . We observe that  $\pi_\phi^k(k_1, \dots, k_d)$  is nonzero when all of  $k_1, \dots, k_r$  are  $\geq k$ , while at least one of them is equal to  $k$ . Consider  $\mathcal{M}_\phi^{k,n}$  for  $k < \infty$ . As in Section 7.1, we break up the sum in its definition according to how many of  $k_1, \dots, k_d$  are equal to  $k$ .

Let  $\mathcal{J}$  be a subset of  $S = [r]$  of size  $|\mathcal{J}| = s \leq r$ . Let  $\mathcal{L} := S \setminus \mathcal{J}$  and  $S^c := [d] \setminus S$ . Note that  $\{\mathcal{J}, \mathcal{L}, S^c\}$  form a partition of the index set  $[d]$ . To simplify notation, let  $\mathcal{J}^c = [d] \setminus \mathcal{J}$  and note that  $\mathcal{J}^c = \mathcal{L} \cup S^c$ .

Consider the terms in the sum (28) for which  $k_j = k$  for  $j \in \mathcal{J}$  and  $k_j > k$  for  $j \in \mathcal{L}$ . We call the sum over these terms  $T_{\mathcal{J}}$ . Then,  $\mathcal{M}_\phi^{k,n} = \sum_{s=1}^r \sum_{\mathcal{J}: |\mathcal{J}|=s} T_{\mathcal{J}}$ .

Now fix some  $s \in [r]$  and some  $\mathcal{J} \subset S$  with  $|\mathcal{J}| = s$ . The  $R$ -terms in the expression of  $T_{\mathcal{J}}$  corresponding to nodes are easy to deal with. For the  $R$ -terms corresponding to edges, we first break them into three categories, based on how many of the endpoints are in  $\mathcal{J}$  (i.e.,  $|e \cap \mathcal{J}| = 0, 1, 2$ ). The case where exactly one endpoint is in  $\mathcal{J}$  (i.e.,  $|e \cap \mathcal{J}| = 1$ ) is further broken into two cases based on whether the other endpoint is in  $\mathcal{L}$  or in  $S^c$ . The former case, i.e.  $|e \cap \mathcal{J}| = |e \cap \mathcal{L}| = 1$  behaves the same as the case  $|e \cap \mathcal{J}| = 2$ . We thus combine these two cases, denoted as  $|e \cap \mathcal{J}| \geq 1, e \subset S$ . To summarize, we break the edges into a total of three categories. We get the following decomposition

$$T_{\mathcal{J}} = \prod_{j \in \mathcal{J}} R_k^n\{j\} \prod_{\substack{|e \cap \mathcal{J}| \geq 1, \\ e \subset S}} R_k^n(e) \times \sum_{\substack{k_j > k, j \in \mathcal{L} \\ k_j \geq 1, j \in S^c}} \left\{ A e^{-\sum_{j \in \mathcal{J}^c} \beta_j k_j} \underbrace{\prod_{j \in \mathcal{J}^c} R_{k_j}^n\{j\}}_{(*)} \underbrace{\prod_{\substack{|e \cap \mathcal{J}|=1, \\ e \cap S^c = \{\ell\}}} R_{k \wedge k_\ell}^n(e)}_{(**)} \underbrace{\prod_{|e \cap \mathcal{J}|=0} R_{k_e}^n(e)}_{(***)} \right\} \quad (39)$$

As in Sections 7.1 and 7.2, we can apply Lemma 6 to decouple the sum and obtain an upper bound on  $T_{\mathcal{J}}$ . The products denoted by  $(*)$ ,  $(**)$  and  $(***)$  produce  $S$ ,  $L$ , and  $M$ -terms<sup>1</sup>,

<sup>1</sup>Strictly speaking, some of the terms produced by  $(***)$  will have the form of an  $M$ -term in the extended sense to be introduced in (44). For example, we will have  $M$ -terms of the form  $M_{1,k+1}^{\infty,\infty}(e)$ . Since every term of the sum is nonnegative, we have the inequality  $M_{k+1}^{\infty,\infty}(e) \leq M_{1,k+1}^{\infty,\infty}(e) \leq M_1^{\infty}(e)$ , which in view of Theorem 2 implies  $\frac{1}{n} \log M_{1,k+1}^{\infty,\infty}(e) \stackrel{\varepsilon}{\asymp} I_e$ .



respectively. Using the same lower bounding technique and applying Theorem 2, we obtain

$$\begin{aligned} \frac{1}{n} \log T_J &\stackrel{\varepsilon}{\asymp} \sum_{j \in \mathcal{J}} I_j + \sum_{\substack{|e \cap \mathcal{J}| \geq 1, \\ e \subset \mathcal{S}}} I_e + \sum_{j \in \mathcal{J}^c} I_j + \sum_{\substack{|e \cap \mathcal{J}|=1, \\ |e \cap \mathcal{S}^c|=1}} I_e + \sum_{|e \cap \mathcal{J}|=0} I_e \\ &= \sum_{j \in V} I_j + \sum_{e \in E} I_e. \end{aligned}$$

Since this expression does not depend on  $\mathcal{J}$ , using log-sum-max rule (R7) as before, we obtain that  $\frac{1}{n} \log \mathcal{M}_\phi^{k,n} \stackrel{\varepsilon}{\asymp} \sum_{j \in V} I_j + \sum_{e \in E} I_e$ .

Now, we need to analyze  $\mathcal{M}_\phi^{\infty,n}$ . We try to break up the sum as before into  $\tilde{T}_J$  terms (defined similar to  $T_J$  for  $\mathcal{M}_\phi^{k,n}$ ). This time however, we only need to consider  $\mathcal{J} = \mathcal{S}$  (and  $\mathcal{L}$  the empty set), because  $\phi = \infty$  implies  $\lambda_j = \infty$  for all  $j \in \mathcal{S}$ . The expansion for  $\tilde{T}_\mathcal{S}$  can be obtained from (39) by setting  $k = \infty$  and removing the terms corresponding to indices in  $\mathcal{S} = \mathcal{J} \cup \mathcal{L}$ ,

$$\mathcal{M}_\phi^{\infty,n} = \tilde{T}_\mathcal{S} = \sum_{k_j \geq 1, j \in \mathcal{S}^c} \left\{ A e^{-\sum_{j \in \mathcal{S}^c} \beta_j k_j} \prod_{j \in \mathcal{S}^c} R_{k_j}^n\{j\} \prod_{\substack{|e \cap \mathcal{S}|=1, \\ e \cap \mathcal{S}^c = \{\ell\}}} R_{k_\ell}^n(e) \prod_{|e \cap \mathcal{S}|=0} R_{k_e}^n(e) \right\}.$$

It follows that

$$\frac{1}{n} \log \mathcal{M}_\phi^{\infty,n} \stackrel{\varepsilon}{\asymp} \sum_{j \in \mathcal{S}^c} I_j + \sum_{\substack{|e \cap \mathcal{S}|=1, \\ |e \cap \mathcal{S}^c|=1}} I_e + \sum_{|e \cap \mathcal{S}|=0} I_e.$$

The last two sums can be described as the sum over all edges  $e : e \cap \mathcal{S}^c \neq \emptyset$ . Putting the pieces together, we have

$$\begin{aligned} \frac{1}{n} \log D_\phi^{k,n} &\stackrel{\varepsilon}{\asymp} \left( \sum_{j \in V} I_j + \sum_{e \in E} I_e \right) - \left( \sum_{j \in \mathcal{S}^c} I_j + \sum_{e \cap \mathcal{S}^c \neq \emptyset} I_e \right) \\ &= \sum_{j \in \mathcal{S}} I_j + \sum_{e \subset \mathcal{S}} I_e \end{aligned}$$

as desired.

## 8 Proof of Theorem 2

Let us start by understanding the asymptotic behavior of  $\frac{1}{n} \log R_u^n(e)$ . Throughout, we fix  $e \in \tilde{E}$ . We either have  $e = \{j\}$  in which case  $\lambda_e = \lambda_j$ , or  $e = \{i, j\}$  in which case  $\lambda_e = \lambda_i \wedge \lambda_j$ . Recall that  $m_* = (m_1, \dots, m_d) \in \mathbb{N}^d$  is a multi-index, and we will work under the collection  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$  of conditional distributions (see Definition 1 for details). The same convention is used regarding the meaning of  $m_e$ , that is,  $m_e = m_j$  for  $e = j$ , and  $m_e = m_i \wedge m_j$  for  $e = \{i, j\}$ . We also fix some  $k \in \mathbb{N}$ , which is the parameter  $k$  appearing in Definition 1 (reserved for the ultimate conditioning on  $\{\phi = k\}$ ). Finally, we always assume  $\varepsilon \in (0, 1)$ .

At first, we need to be careful about whether  $u < m_e$  or  $u \geq m_e$ .

**Lemma 7.** *Let  $u \in [n]$  and assume  $u \geq m_e$ . Then,*

$$\mathbb{P}_{\lambda_*}^{m_*} \left( \left| \frac{1}{n-u+1} \log R_u^n(e) - I_e \right| > \varepsilon \right) \leq 2 \exp \left[ -\frac{(n-u+1)\varepsilon^2}{2M^2} \right] \quad (40)$$

*Proof.* Since  $m_e \leq u$ , conditioned on  $\mathbb{P}_{\lambda_*}^{m_*}$ ,  $\mathbf{X}_e^u, \mathbf{X}_e^{u+1}, \dots$  are i.i.d. from  $f_e$ . Recalling definition (22),  $\log R_u^n(e) = \sum_{t=u}^n h_e(\mathbf{X}_e^t)$  which is a sum of  $(n-u+1)$  i.i.d. bounded variables  $h_e(\mathbf{X}_e^t) \in [-M, M]$  with mean  $\mathbb{E}_{f_e} h_e(\mathbf{X}_e^u) = I_e$ . The result then follows from Hoeffding inequality.  $\square$

Before moving on, we need an extension of Definition 1. We need to deal with intermediate sequences whose terms depend possibly on  $m_*$  (in addition to  $k$ ). There is nothing to preclude such dependence in Definition 1. Hence, we use the same definition for  $\varepsilon$ -equivalence of such sequences with respect to the collection  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ . Note that for any  $u, \nu \in \mathbb{N}$ , we can write

$$R_u^n(e) = R_u^{\nu-1}(e) R_{u \vee \nu}^n(e) \quad (41)$$

which holds irrespective of whether  $u \geq \nu$  or  $u < \nu$ .

**Lemma 8.** *For any  $u \in [k]$ ,  $\frac{1}{n} \log R_{u \vee m_e}^n(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  with respect to  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$*

**Lemma 9.** *For any  $u \in \mathbb{N}$ ,  $\frac{1}{n} \log R_u^{m_e-1}(e) \stackrel{\varepsilon}{\asymp} 0$  as  $n \rightarrow \infty$  with respect to  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .*

**Lemma 10.** *For any  $u \in [k]$ ,  $\frac{1}{n} \log R_u^n(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  with respect to  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .*

The last lemma proves the statement in Theorem 2 regarding asymptotic behavior of  $\frac{1}{n} \log R_u^n(e)$  for  $u \in [k]$ .

*Proof of Lemma 8.* Apply Lemma 7 with  $u$  replaced with  $u \vee m_e$ . Since  $\varepsilon < 1$  and  $u \vee m_e \leq k \vee m_e$ , the RHS of (40) is further bounded above by

$$2 \exp \left( \frac{(k \vee m_e)\varepsilon}{2M^2} \right) \exp \left( -\frac{n\varepsilon^2}{2M^2} \right) \leq 2 \exp \left( -\frac{n\varepsilon^2}{4M^2} \right)$$

as long as  $(k \vee m_e)\varepsilon \leq n\varepsilon^2/2$  or equivalently  $n\varepsilon \geq 2(k \vee m_e)$ . (This same condition guarantees  $u \vee m_e \in [n]$  justifying application of Lemma 7.) The condition obtained is of the form required by Definition 1, since  $2(k \vee m_e)$  is bounded above by a polynomial, say  $2(k + m_i)$  if  $e = \{i, j\}$ . This shows that

$$\frac{1}{n-u \vee m_e+1} \log R_{u \vee m_e}^n(e) \stackrel{\varepsilon}{\asymp} I_e \quad \text{w.r.t. } \{\mathbb{P}_{\lambda_*}^{m_*}\}.$$

Now, note that  $|\frac{n-u \vee m_e+1}{n} - 1| \leq \frac{u \vee m_e-1}{n} \leq \frac{k \vee m_e}{n}$  which can be made  $\leq \varepsilon$  by choosing  $n\varepsilon \geq (k \vee m_e)$ . This implies that  $\frac{n-u \vee m_e+1}{n} \stackrel{\varepsilon}{\asymp} 1$ . Applying rule (R5), with  $a_n = \frac{1}{n-u \vee m_e+1} \log R_{u \vee m_e}^n(e)$ ,  $b_n = I_e$  and  $c_n = \frac{n-u \vee m_e+1}{n}$ , we obtain the desired result.  $\square$

*Proof of Lemma 9.* If  $u > m_e - 1$ , we have by definition  $R_u^{m_e-1}(e) = 1$  and there is nothing to show. Otherwise, by boundedness assumption  $\|h\|_\infty \leq M$ , we have

$$e^{-Mm_e} \leq e^{-M(m_e-u)} \leq R_u^{m_e-1}(e) \leq e^{M(m_e-u)} \leq e^{Mm_e}.$$

Hence, by taking  $n\varepsilon \geq Mm_e$ , we have  $|\frac{1}{n} \log R_u^{m_e-1}(e)| \leq \varepsilon$ , which implies the result.  $\square$

*Proof of Lemma 10.* Apply (41) with  $\nu = m_e$ , to obtain

$$\frac{1}{n} \log R_u^n(e) = \frac{1}{n} \log R_u^{m_e-1}(e) + \frac{1}{n} \log R_{u \vee m_e}^n(e).$$

The result now follows from Lemmas 8 and 9 and rule (R3).  $\square$

### 8.1 Bounding $S$ -terms

Bounding  $S$ -terms is perhaps the most elaborate part of the proof. We start with a uniformization of Lemma 7 and then proceed in steps, working on various parts of the sum  $S_u^\infty(e) := S_u^{\infty, n}(e)$  one at a time. Up to Lemma 16, we will use the shorthand notation introduced in (27), with  $n$  superscript dropped. It might help to recall that in this notation,  $u$  and  $\infty$  are the initial and final indices of the sum, respectively. Also, the edge  $e \in E$  is fixed throughout.

**Lemma 11.** *Let  $u \in \mathbb{N}$  and  $\alpha \in (0, 1)$  such that  $m_e \leq u \leq \lfloor \alpha n \rfloor$ . Then,*

$$\sup_{u \leq p \leq \lfloor \alpha n \rfloor} \left| \frac{1}{n-p+1} \log R_p^n(e) - I_e \right| \leq \varepsilon$$

*with  $\mathbb{P}_{\lambda_*}^{m_*}$ -probability at least  $1 - 2(\lfloor \alpha n \rfloor - u + 1) \exp[-\frac{\varepsilon^2((1-\alpha)n+1)}{2M^2}]$ .*

**Lemma 12.** *Let  $u \in [k]$  and  $\alpha \in (0, 1)$  such that  $m_e \leq u \leq \lfloor \alpha n \rfloor$ . Then*

$$\mathbb{P}_{\lambda_*}^{m_*} \left( \left| \frac{1}{n} \log S_u^{\lfloor \alpha n \rfloor}(e) - I_e \right| \leq 2\varepsilon \right) \geq 1 - 2n \exp \left( -\frac{1-\alpha}{2M^2} n \varepsilon^2 \right)$$

*for  $n\varepsilon \geq c_0 k$  and  $\varepsilon \in (0, 1)$ .*

**Lemma 13.** *Let  $\delta \in (0, 1)$  and  $\alpha = \frac{M+\delta\beta}{M+\beta}$ . Then, for  $n \geq n_0(A, \beta, M, \delta)$ ,*

$$\frac{1}{n} \log S_{\lfloor \alpha n \rfloor + 1}^n(e) \leq -\frac{\delta}{2} \beta.$$

**Lemma 14.** *Let  $\alpha = \frac{M+\frac{1}{2}\beta}{M+\beta}$  and  $u \in [k]$ . Then,  $\frac{1}{n} \log S_{u \vee m_e}^{\lfloor \alpha n \rfloor}(e) \overset{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .*

**Lemma 15.** *For any  $u \in [k]$ , we have  $\frac{1}{n} \log S_{u \vee m_e}^n(e) \overset{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .*

*Proof of Lemma 11.* We note that for any  $p = u, u+1, \dots, \lfloor \alpha n \rfloor$ , Lemma 7 applies. We can further upper-bound the RHS of (40) by

$$2 \exp \left[ -\frac{(n-p+1)\varepsilon^2}{2M^2} \right] \leq 2 \exp \left[ -\frac{(n-\alpha n+1)\varepsilon^2}{2M^2} \right].$$

The result follows by applying union bound.  $\square$

*Proof of Lemma 12.* By Lemma 11, uniformly over  $p = u, \dots, \lfloor \alpha n \rfloor$ , we have

$$e^{(n-p+1)(I_e-\varepsilon)} \leq R_p^n(e) \leq e^{(n-p+1)(I_e+\varepsilon)} \quad (42)$$

with  $\mathbb{P}_{\lambda_*}^{m*}$ -probability at least  $1 - 2n \exp(-\frac{(1-\alpha)n\varepsilon^2}{2M^2})$ . (Note that this is a further lower bound w.r.t. that of Lemma 11) On the event that (42) holds, we have

$$\begin{aligned} S_u^{[\alpha n]}(e) &\leq \sum_{p=u}^{[\alpha n]} A e^{-\beta p} e^{(n-p+1)(I_e + \varepsilon)} \\ &\leq A e^{(n+1)(I_e + \varepsilon)} \sum_{p=1}^{\infty} e^{-(\beta + I_e)p} = \frac{A e^{(n+1)(I_e + \varepsilon)}}{e^{\beta + I_e} - 1}. \end{aligned}$$

Take  $C_1 := \max\{0, \log \frac{A}{e^{\beta + I_e} - 1}\}$ . Then,

$$\frac{1}{n} \log S_u^{[\alpha n]}(e) \leq \frac{C_1}{n} + \frac{n+1}{n} (I_e + \varepsilon) \leq I_e + 2\varepsilon$$

as long as  $n\varepsilon \geq C_1 + I_e + 1$  (and  $\varepsilon < 1$ ). To get the lower bound, we note that (42) implies

$$\begin{aligned} S_u^{[\alpha n]}(e) &\geq \sum_{p=u}^{[\alpha n]} A e^{-\beta p} e^{(n-p+1)(I_e - \varepsilon)} \\ &\geq A e^{(n+1)(I_e - \varepsilon)} e^{-(\beta + I_e)u} \end{aligned}$$

where we have lower bounded a sum of nonnegative terms by its first term. Hence,

$$\begin{aligned} \frac{1}{n} \log S_u^{[\alpha n]}(e) &\geq \frac{n+1}{n} (I_e - \varepsilon) - \frac{|\log A| + (\beta + I_e)u}{n} \\ &\geq I_e - \varepsilon - \frac{1 + |\log A| + (\beta + I_e)k}{n} \geq I_e - 2\varepsilon \end{aligned}$$

as long as  $n\varepsilon \geq 1 + |\log A| + (\beta + I_e)k$ .  $\square$

*Proof of Lemma 13.* By boundedness assumption  $\|h\|_{\infty} \leq M$ , we have  $R_p^n(e) \leq e^{(n-p+1)M}$  as long as  $p \leq n$ . Hence,

$$\begin{aligned} S_{[\alpha n]+1}^n(e) &\leq \sum_{p=[\alpha n]+1}^n A e^{-\beta p} e^{(n-p+1)M} \\ &= A e^{(n+1)M} \sum_{p=[\alpha n]+1}^n e^{-(\beta + M)p} \\ &\leq A e^{(n+1)M} (n - \alpha n + 1) e^{-(\beta + M)\alpha n} \end{aligned}$$

where we have used  $[\alpha n] > \alpha n - 1$ . Taking  $\alpha$  to be as stated and noting that  $1 - \alpha \in (0, 1)$ , we get

$$\frac{1}{n} S_{[\alpha n]+1}^n(e) \leq \frac{|\log A|}{n} + \frac{n+1}{n} M + \frac{\log((1-\alpha)n+1)}{n} - M - \delta\beta \leq -\frac{\delta}{2}\beta$$

as long as  $n \geq n_0(A, \beta, M, \delta)$  for some  $n_0$  large enough.  $\square$

*Proof of Lemma 14.* Apply Lemma 12 with  $u$  replaced with  $u \vee m_e$ . To ensure  $u \vee m_e \leq \lfloor \alpha n \rfloor$ , let  $n \geq \frac{1}{\alpha}(k \vee m_e + 1)$ . To ensure that the bound of Lemma 12 holds, let  $n\varepsilon \geq c_0 k$ . Since, these two conditions are met if  $n\varepsilon \geq \frac{1}{\alpha}(k + m_e) + c_0 k$ , the result follows. (Note also that  $\frac{1-\alpha}{2M^2}$  is a positive constant by our choice of  $\alpha$ .)  $\square$

*Proof of Lemma 15.* Let  $\alpha = \frac{M+\frac{1}{2}\beta}{M+\beta}$  and as in Lemma 14 assume  $n \geq \frac{1}{\alpha}(k \vee m_e + 1)$  so that  $u \vee m_e \leq \lfloor \alpha n \rfloor$ . (This is just to make sure that sums ranging from  $u \vee m_e$  to  $\lfloor \alpha n \rfloor$  are not vacuous.) By Lemma 12, we have

$$\frac{1}{n} \log S_{\lfloor \alpha n \rfloor + 1}^n(e) \stackrel{\varepsilon}{\preceq} -\frac{1}{4}\beta$$

and by Lemma 14,  $\frac{1}{n} \log S_{u \vee m_e}^{\lfloor \alpha n \rfloor}(e) \stackrel{\varepsilon}{\asymp} I_e$ . Now, we can break up the sum and use log-sum-max rule (R7),

$$\begin{aligned} \frac{1}{n} \log S_{u \vee m_e}^n(e) &= \frac{1}{n} \log \left( S_{u \vee m_e}^{\lfloor \alpha n \rfloor}(e) + S_{\lfloor \alpha n \rfloor + 1}^n(e) \right) \\ &\stackrel{\varepsilon}{\asymp} \max \left\{ \underbrace{\frac{1}{n} \log S_{u \vee m_e}^{\lfloor \alpha n \rfloor}(e)}_{\stackrel{\varepsilon}{\asymp} I_e}, \underbrace{\frac{1}{n} \log S_{\lfloor \alpha n \rfloor + 1}^n(e)}_{\stackrel{\varepsilon}{\preceq} -\frac{1}{4}\beta} \right\} \stackrel{\varepsilon}{\asymp} I_e \end{aligned}$$

where the last  $\stackrel{\varepsilon}{\asymp}$  follows from rule (R6).  $\square$

The next step is to move from  $S_{u \vee m_e}^n(e)$  to  $S_u^n(e)$ . We need a couple of lemmas. To simplify notation, throughout this section, let

$$\xi := \xi(u, m_e) := u \vee m_e. \quad (43)$$

We occasionally drop the dependence of  $\xi$  on  $u$  and  $m_e$  (although this is implicitly assumed). We note that all the lemmas established so far in this section hold, if we replace  $[k]$  in their statements with  $[2k]$  (or any other constant multiple of  $k$ ). For the rest of this subsection, we will use the full superscript notation  $S_u^{\nu, n}(e)$  introduced in (23).

**Lemma 16.** For  $1 \neq u \in [2k]$ , we have  $\frac{1}{n} \log S_{u-1}^{\xi-1, \xi-1}(e) \stackrel{\varepsilon}{\asymp} 0$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .

**Lemma 17.** For  $1 \neq u \in [2k]$ , we have  $\frac{1}{n} \log S_{u-1}^{\xi-1, n}(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .

**Lemma 18.** For  $1 \neq u \in [2k]$ , we have  $\frac{1}{n} \log S_{u-1}^{n, n}(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .

**Lemma 19.** For  $u \in [k]$ , we have  $\frac{1}{n} \log S_u^{\infty, n}(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .

The last lemma completes the proof of the statement in Theorem 2 regarding the  $S$  terms.

*Proof of Lemma 16.* For  $p \leq \xi - 1$ ,  $e^{-M(\xi-p)} \leq R_p^{\xi-1}(e) \leq e^{M(\xi-p)}$ , by boundedness assumption. Hence, we have

$$\begin{aligned} S_{u-1}^{\xi-1, \xi-1}(e) &\leq \sum_{p=u-1}^{\xi-1} A e^{-\beta p} e^{M(\xi-p)} \leq \frac{A e^{M\xi}}{1 - e^{-(\beta+M)}}, \\ S_{u-1}^{\xi-1, \xi-1}(e) &\geq \sum_{p=u-1}^{\xi-1} A e^{-\beta p} e^{-M(\xi-p)} \geq A e^{-\beta(\xi-1)} e^{-M} \end{aligned}$$

Let  $C_1 = |\log \frac{A}{1-e^{-(\beta+M)}}|$  and  $C_2 = |\log(Ae^{\beta-M})|$ . We have

$$-\frac{C_2}{n} - \frac{2\beta(k \vee m_e)}{n} \leq \frac{1}{n} \log S_{u-1}^{\xi-1, \xi-1}(e) \leq \frac{C_1}{n} + \frac{2M(k \vee m_e)}{n}$$

where we have used  $\xi \leq (2k) \vee m_e$  which follows from definition 43 and assumption  $u \in [2k]$ . It follows that  $|\frac{1}{n} \log S_{u-1}^{\xi-1, \xi-1}(e)| \leq \varepsilon$  if we take  $n\varepsilon \geq C_3(k + m_e)$ , proving the result.  $\square$

*Proof of Lemma 17.* For any  $p \in \{u-1, \dots, \xi-1\}$ , we have by (41),

$$R_p^n(e) = R_p^{\xi-1}(e) R_\xi^n(e).$$

It follows from the definition of  $S$  term that

$$S_{u-1}^{\xi-1, n}(e) = R_\xi^n(e) \sum_{p=u-1}^{\xi-1} A e^{-\beta p} R_p^{\xi-1}(e) = R_\xi^n(e) S_{u-1}^{\xi-1, \xi-1}(e).$$

The result now follows from Lemmas 8 and 16.  $\square$

*Proof of Lemma 18.* We have  $S_{u-1}^{n, n}(e) = S_{u-1}^{\xi-1, n}(e) + S_\xi^{n, n}(e)$ . The result now follows from Lemmas 17 and 15, and log-sum-max rule (R7).  $\square$

Note that since  $[k+1] \subset [2k]$ , it follows that  $\frac{1}{n} \log S_u^{n, n}(e) \stackrel{\varepsilon}{\asymp} I_e$  for all  $u \in [k]$ . The final step is to move from  $S_u^{n, n}(e)$  to  $S_u^{\infty, n}(e)$ .

*Proof of Lemma 19.* We have  $S_u^{\infty, n}(e) = S_u^{n, n}(e) + S_{n+1}^{\infty, n}(e)$ . Since  $R_p^n(e) = 1$  for all  $p > n$  (by convention), we have

$$S_{n+1}^{\infty, n}(e) = \sum_{p=n+1}^{\infty} A e^{-\beta p} = \frac{A e^{-\beta(n+1)}}{1 - e^{-\beta}}.$$

It follows that  $\frac{1}{n} \log S_{n+1}^{\infty, n}(e) \stackrel{\varepsilon}{\asymp} -\beta$ . Then, by rules (R7) and (R4),

$$\begin{aligned} \frac{1}{n} \log S_u^{\infty, n}(e) &\stackrel{\varepsilon}{\asymp} \max \left\{ \frac{1}{n} \log S_u^{n, n}(e), \frac{1}{n} \log S_{n+1}^{\infty, n}(e) \right\} \\ &\stackrel{\varepsilon}{\asymp} \max \{I_e, -\beta\} = I_e \end{aligned}$$

where we have used Lemma 18.  $\square$

## 8.2 Bounding $M$ -terms

With some work, we can reduce bounding  $M$ -terms to that of bounding  $R$  and  $S$ -terms.

**Lemma 20.** For  $u \in [k]$ , we have  $M_u^n(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t  $\{\mathbb{P}_{\lambda_*}^{m*}\}$ .

*Proof.* Let  $n \geq k$ , so that the sums are not vacuous. For  $q \in \{u, \dots, n\}$  the cardinality of the set  $\{(p_1, p_2) : u \leq p_1, p_2 \leq n, p_1 \wedge p_2 = q\}$  is  $2(n - q) + 1$ . Hence,

$$\begin{aligned} M_u^n(e) &\leq \sum_{p_1=u}^n \sum_{p_2=u}^n A e^{-(\beta_1 \wedge \beta_2)(p_1 \wedge p_2)} R_{p_1 \wedge p_2}^n(e) \\ &= A \sum_{q=u}^n [2(n - q) + 1] e^{-(\beta_1 \wedge \beta_2)q} R_q^n(e) \\ &\leq n \sum_{q=u}^n 2A e^{-(\beta_1 \wedge \beta_2)q} R_q^n(e). \end{aligned}$$

Note that this last sum is of the form  $S_u^n(e)$ . For the lower bound, we use the first term of the sum,  $M_u^n(e) \geq A e^{-(\beta_1 + \beta_2)u} R_u^n(e)$ . Since  $u \leq k$ , we have

$$-\frac{|\log A| + (\beta_1 + \beta_2)k}{n} + \frac{1}{n} \log R_u^n(e) \leq \frac{1}{n} \log M_u^n(e) \leq \frac{\log n}{n} + \frac{1}{n} \log S_u^n(e).$$

The only new term (with respect to what established earlier) is  $\log n/n$  which is  $\stackrel{\varepsilon}{\asymp} 0$ . This can be seen by noting that  $|\log n/n| \leq \varepsilon$  if  $\sqrt{n}\varepsilon \geq 1$ . The result now follows from Lemmas 10 and 19.  $\square$

To move from  $M_u^n(e)$  to  $M_u^\infty(e)$ , we introduce the following extended notation

$$M_{a,b}^{c,d}(e) := M_{a,b}^{c,d,n}(e) := \sum_{p_1=a}^c \sum_{p_2=b}^d A e^{-(\beta_1 p_1 + \beta_2 p_2)} R_{p_1 \wedge p_2}^n(e) \quad (44)$$

so that  $M_u^\infty(e) = M_{u,u}^{\infty,\infty}(e)$ .

**Lemma 21.** *For  $u \in [k]$ , we have  $M_u^\infty(e) \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m*}\}$ .*

*Proof.* Let  $n \geq k$ . The strategy is to break up the sum as

$$M_{u,u}^{\infty,\infty}(e) = M_{u,u}^{n,n}(e) + M_{n+1,u}^{\infty,n}(e) + M_{u,n+1}^{n,\infty}(e) + M_{n+1,n+1}^{\infty,\infty}(e) \quad (45)$$

and then apply the log-sum-max rule (R7). The first term is taken care of by Lemma 20. For the second term, we have

$$\begin{aligned} M_{n+1,u}^{\infty,n}(e) &= \sum_{p_1=n+1}^{\infty} \sum_{p_2=u}^n A e^{-(\beta_1 p_1 + \beta_2 p_2)} R_{p_2}^n(e) \\ &= C_1 e^{-\beta_1(n+1)} S_u^n(e) \end{aligned}$$

Applying Lemma 18 we get  $\frac{1}{n} \log M_{n+1,u}^{\infty,n}(e) \stackrel{\varepsilon}{\asymp} -\beta_1 + I_e$ . The third term in 45 is similar. Recalling that  $R_p^n(e) = 1$  for  $p > n$ , the fourth term,  $M_{n+1,n+1}^{\infty,\infty}(e)$ , is equal to  $C_2 e^{-(\beta_1 + \beta_2)(n+1)}$ . Hence, by (R7),

$$\frac{1}{n} \log M_{u,u}^{\infty,\infty}(e) \stackrel{\varepsilon}{\asymp} \max\{I_e, -\beta_1 + I_e, -\beta_2 + I_e, -\beta_1 - \beta_2\} = I_e.$$

$\square$

### 8.3 Bounding $L$ -terms

**Lemma 22.** *For  $u, r \in [k]$ , we have  $\frac{1}{n} \log L_{u,(r)}^\infty \stackrel{\varepsilon}{\asymp} I_e$  as  $n \rightarrow \infty$  w.r.t.  $\{\mathbb{P}_{\lambda_*}^{m_*}\}$ .*

*Proof.* First consider the case  $u \geq r$ . Then, we have

$$L_{u,(r)}^\infty = \sum_{p=u}^{\infty} A e^{-\beta p} R_r^n(e) = C_1 e^{-\beta u} R_r^n(e).$$

Since  $|\frac{\beta u}{n}| \leq \frac{\beta k}{n}$ , we have  $\frac{1}{n} \log(C_1 e^{-\beta u}) \stackrel{\varepsilon}{\asymp} 0$ . The result now follows from Lemma 10. For the case  $u < r$ , we have

$$\begin{aligned} L_{u,(r)}^\infty &= \sum_{p=u}^{r-1} A e^{-\beta p} R_p^n(e) + \sum_{p=r}^{\infty} A e^{-\beta p} R_r^n(e) \\ &= S_u^{r-1}(e) + C_1 e^{-\beta r} R_r^n(e). \end{aligned} \tag{46}$$

Let  $n \geq k$  so that  $n \geq r$ . Note that  $A e^{-u\beta} R_u^n(e) \leq S_u^{r-1}(e) \leq S_u^n(e)$ . It follows from Lemmas 18 and 10, and  $\frac{1}{n} \log(A e^{-\beta u}) \stackrel{\varepsilon}{\asymp} 0$  that  $\frac{1}{n} \log S_u^{r-1}(e) \stackrel{\varepsilon}{\asymp} I_e$ . Applying rule (R7) to (46) and using a similar argument for the second term, we get the result.  $\square$

## 9 Conclusion

We have introduced a graphical model framework which allows for modeling and detection of multiple change points in networks. Within this framework, we proposed stopping rules for the detection of change points and particular functionals of them (the minimum over a subset), based on thresholding the posterior probabilities. A message passing algorithm for efficient computation of these posteriors was derived. It was also shown that the proposed rules are asymptotically optimal in terms of their expected delay, within the Bayesian framework.

Let us discuss some directions for possible extension of this work. The assumption that the distribution of shared (edge) information between two nodes only depends on the minimum of the associated change points (cf. discussion after equation (3)) might be restrictive in practice. The current assumption simplifies the analysis in many places and it has an impact on the asymptotic delay. For example, we suspect that the “no gain” phenomenon in asymptotic delay for detection of a single change point, discussed in Remark 3 after Theorem 1, is due to this rather simplistic assumption. It will be interesting to be able to extend the analysis to a model which allows for a more general dependence on the two change points. At present, however, we do not know how much of our analysis can be carried over to the general case.

It is possible to derive an approximate message passing algorithm with computational cost scaling as  $O(|V| + |E|)$  for each time step  $n$ . That is, the computational cost is constant in time  $n$ . Simulations indicate that this fast algorithm approximates the exact message passing well. The presentation of the algorithm and its theoretical analysis will be deferred to a future publication.

As was discussed in the remarks after Theorems 1 and 2, the assumptions on the likelihood ratio, i.e., the boundedness, and the priors, i.e., exponential tail decay are crucial to our proof. They seem to strike the right balance between the prior and the likelihood and they also allow for the break-up of the analysis of the rather complicated likelihood ratios (cf. (28)) into



simpler pieces. This is in contrast to the more classical case of a single change point where the analysis goes through seamlessly, say, irrespective of the tail behavior of the priors [18]. Whether these limitations are genuinely present in the multiple change point model or are artifacts of the proof technique is not clear at this point.

Finally, although our main focus in this paper was on the Bayesian formulation, we note that there are non-Bayesian optimality criteria for the single-change point problem, e.g., the minimax as considered in [19]. It is an interesting question whether one can derive minimax optimal rules for the model we consider here.

## A Proof of Lemma 2

Consider, for example, node  $i_1$  and let  $j$  be one of its neighbors in  $G$ , i.e.  $\{i_1, j\} \in E$ . Let  $k_1 \geq n + 1$ . Then  $P(\mathbf{X}_{i_1}^n | \lambda_{i_1} = k_1) = \prod_{t=1}^n g_{i_1}(X_{i_1}^t) = P(\mathbf{X}_{i_1}^n | \lambda_{i_1} = n + 1)$ . Similarly, the distribution of  $X_{i_1 j}$  given  $\lambda_{i_1} = k_1$  and  $\lambda_j$  is independent of the particular value of  $k_1$ , that is,

$$P(\mathbf{X}_{i_1 j}^n | \lambda_{i_1} = k_1, \lambda_j) = P(\mathbf{X}_{i_1 j}^n | (n + 1) \wedge \lambda_j) = P(\mathbf{X}_{i_1 j}^n | \lambda_{i_1} = n + 1, \lambda_j).$$

Let  $\mathcal{J} := \{i_1, \dots, i_r\}$  and  $\mathcal{J}^c = [d] \setminus \mathcal{J}$ . Pick  $k_j \geq n + 1$  for  $j \in \mathcal{J}$ . Then, the argument above applied to each node in  $\mathcal{J}$  shows that

$$\begin{aligned} P(\mathbf{X}_*^n | \lambda_j = k_j, j \in \mathcal{J}) &= \sum_{\lambda_*} P(\mathbf{X}_*^n | \lambda_*) P(\lambda_* | \lambda_j = k_j, j \in \mathcal{J}) \\ &= \sum_{\lambda_\ell, \ell \in \mathcal{J}^c} P(\mathbf{X}_*^n | \lambda_\ell, \ell \in \mathcal{J}^c, \lambda_j = k_j, j \in \mathcal{J}) P(\lambda_\ell, \ell \in \mathcal{J}^c) \\ &= \sum_{\lambda_\ell, \ell \in \mathcal{J}^c} P(\mathbf{X}_*^n | \lambda_\ell, \ell \in \mathcal{J}^c, \lambda_j = n + 1, j \in \mathcal{J}) P(\lambda_\ell, \ell \in \mathcal{J}^c) \end{aligned}$$

where the second inequality follows by independence of  $\{\lambda_i\}$  a priori. As the last expression does not depend on  $\{k_j\}$ , the proof is complete.

## B Proof of Lemma 5

Let  $k_* = (k_1, \dots, k_d) \in \mathbb{N}^d$  be a multi-index. We have

$$\begin{aligned} P(\mathbf{X}_*^n | \phi = k) &= \sum_{k_* \in \mathbb{N}^d} P(\mathbf{X}_*^n | \lambda_* = k_*) \mathbb{P}(\lambda_* = k_* | \phi = k) \\ &= \sum_{k_* \in \mathbb{N}^d} \left\{ \prod_{e \in \tilde{E}} P(\mathbf{X}_e^n | \lambda_e = k_e) \right\} \pi_\phi^k(k_*) \end{aligned}$$

where we have used the extended edge notation of Section 4 and conditional distribution introduced in (16). Using the pre- and post-change densities, we get

$$P(\mathbf{X}_*^n | \phi = k) = \sum_{k_* \in \mathbb{N}^d} \left\{ \prod_{e \in \tilde{E}} \left[ \prod_{t=1}^{k_e-1} g_e(X_e^t) \prod_{t=k_e}^n f_e(X_e^t) \right] \right\} \pi_\phi^k(k_*) \quad (47)$$

where by convention, empty products are equal to 1. Dividing (47) by

$$U(\mathbf{X}_*^n) := \prod_{e \in \tilde{E}} \left[ \prod_{t=1}^n g_e(X_e^t) \right].$$

we obtain

$$\frac{P(\mathbf{X}_*^n \mid \phi = k)}{U(\mathbf{X}_*^n)} = \sum_{k_* \in \mathbb{N}^d} \left\{ \prod_{e \in \tilde{E}} R_{k_e}^n(\mathbf{X}_e) \right\} \pi_\phi^k(k_*) = \mathcal{M}_\phi^{k,n}$$

where we have used definitions (22) and (28). The same expression holds, if we replace  $k$  with  $\infty$ . The result now follows from definition (17) of  $D_\phi^{k,n}$ .

## C Proof of Lemma 6

The idea of the proof is to write the sum as the diagonal part of a higher dimensional one and then drop the restriction to the diagonal. Let us illustrate the idea first by proving (30).

We can write

$$\begin{aligned} \sum_{p \in S_1} e^{-\beta_1 p} F_1(p) H_1(p) &= \sum_{p \in S_1} \sum_{q \in S_1} 1\{p = q\} e^{-\frac{\beta_1}{2}(p+q)} F_1(p) H_1(q) \\ &\leq \sum_{p \in S_1} \sum_{q \in S_1} e^{-\frac{\beta_1}{2}(p+q)} F_1(p) H_1(q) \end{aligned}$$

The bound holds since the terms are nonnegative. Now, the RHS factors over  $p$  and  $q$  and we get (30).

The idea for the proof of (31) is similar. For every pair  $(i, j) \in \binom{[r]}{2}$ , we introduce new versions of  $k_i$  and  $k_j$  so that the corresponding term  $G_{ij}(k_i, k_j)$  involves the new variables. To be more precise, let  $\mathcal{K} = \{\nu_1, \dots, \nu_K\}$  be an enumeration of the elements of  $\binom{[r]}{2}$ . To each element  $\nu_\ell = (i, j) \in \mathcal{K}$  with  $i < j$ , associate variables  $u_\ell^1$  and  $u_\ell^2$ , representing newer versions of  $k_i$  and  $k_j$ . In other words,  $u_\ell^1$  is the new version of  $k_{\nu_\ell(1)}$ .

This procedure introduces  $2K$  extra variables. To each of the original  $k_i$  variables, there corresponds exactly  $r - 1$  new versions. Letting

$$T_0 := \sum_{\mathbf{k} \in \mathbf{S}} \left\{ e^{-\beta^T \mathbf{k}} \prod_{j=1}^r F_j(k_j) \prod_{(i,j) \in \binom{[r]}{2}} G_{ij}(k_i, k_j) \right\}$$

denote the LHS of (31), we have

$$\begin{aligned} T_0 &= \sum_{\{k_j\}, \{(u_\ell^1, u_\ell^2)\}} \left\{ 1\{u_\ell^1 = k_{\nu_\ell(1)}, u_\ell^2 = k_{\nu_\ell(2)}, \ell \in [K]\} \times \right. \\ &\quad \left. \exp \left[ - \sum_j \frac{\beta_j}{r} k_j - \sum_{\ell \in [K]} \left( \frac{\beta_{\nu_\ell(1)}}{r} u_\ell^1 + \frac{\beta_{\nu_\ell(2)}}{r} u_\ell^2 \right) \right] \prod_j F_j(k_j) \prod_{\ell \in \mathcal{K}} G_{\nu_\ell}(u_\ell^1, u_\ell^2) \right\} \end{aligned}$$

where the summation is over  $\{k_j \in S_j, u_\ell^1 \in S_{\nu_\ell(1)}, u_\ell^2 \in S_{\nu_\ell(2)}, j \in [r], \ell \in [K]\}$ . Dropping the indicator, we get an upper bound which separates

$$\begin{aligned} T_0 &\leq \sum_{\{k_j\}, \{(u_\ell^1, u_\ell^2)\}} \left\{ e^{-\sum_j \frac{\beta_j}{r} k_j - \sum_{\ell \in [K]} \left( \frac{\beta_{\nu_\ell(1)}}{r} u_\ell^1 + \frac{\beta_{\nu_\ell(2)}}{r} u_\ell^2 \right)} \prod_j F_j(k_j) \prod_{\ell \in \mathcal{K}} G_{\nu_\ell}(u_\ell^1, u_\ell^2) \right\} \\ &= \prod_j \left\{ \sum_{k_j} e^{-\frac{\beta_j}{r} k_j} F_j(k_j) \right\} \prod_{\ell \in \mathcal{K}} \left\{ \sum_{(u_\ell^1, u_\ell^2)} e^{-\left( \frac{\beta_{\nu_\ell(1)}}{r} u_\ell^1 + \frac{\beta_{\nu_\ell(2)}}{r} u_\ell^2 \right)} G_{\nu_\ell}(u_\ell^1, u_\ell^2) \right\} \end{aligned}$$

which is the desired result.

## D Proof of Lemma 4

(a) We start by proving (18). Pick  $n_0$  large enough so that for  $n \geq n_0$ , we have  $q(n) \leq Cn^a$  for some numerical constant  $a \in \mathbb{N}$ . Fix some  $k \in \mathbb{N}$  throughout the proof. For now, fix  $m_* \in \mathbb{N}^d$  such that  $\pi_\phi^k(m_*) > 0$ . Pick  $\varepsilon_n := \sqrt{\frac{\gamma \log n}{c_1 n}} \wedge \varepsilon_0$  for some  $\gamma$  to be determined shortly. Note that  $\sqrt{n} \geq \frac{1}{\varepsilon_n} p(m_*, k)$  is equivalent to  $\sqrt{\frac{\gamma}{c_1} \log n} \wedge (\varepsilon_0 \sqrt{n}) \geq p(m_*, k)$ , which holds for sufficiently large  $n$ . Let  $n_1 := n_1(m_*, k)$  be the smallest  $n$  for which this inequality holds.

Using the shorthand notation  $D_\phi^{k,n} = D_\phi^k(\mathbf{X}_*^n)$ , we have for  $n \geq \max\{n_0, n_1\}$ ,

$$\mathbb{P}_{\lambda_*}^{m_*} \left\{ \left| \frac{1}{n} \log D_\phi^{k,n} - I_\phi \right| > \varepsilon_n \right\} \leq Cn^a \exp(-c_1 n \varepsilon_n^2) = Cn^{a-\gamma}$$

Taking  $\gamma = a + 2$ , we have by Borel-Cantelli lemma that  $\mathbb{P}_{\lambda_*}^{m_*} \left\{ \frac{1}{n} \log D_\phi^{k,n} \xrightarrow{n \rightarrow \infty} I_\phi \right\} = 1$ . It follows that the sequence  $\left\{ \frac{1}{n+k} \log D_\phi^{k,k+n} \right\}_n$  has the same limit a.s.  $\mathbb{P}_{\lambda_*}^{m_*}$ . Since  $k$  is fixed for now,  $\frac{n}{n+k} \sim 1$  as  $n \rightarrow \infty$ , hence  $\mathbb{P}_{\lambda_*}^{m_*} \left\{ \frac{1}{n} \log D_\phi^{k,k+n} \xrightarrow{n \rightarrow \infty} I_\phi \right\} = 1$ .

We now take the average with respect to conditioned distribution of  $\lambda_*$  given  $\phi = k$ . That is, we multiply by  $\pi_\phi^k(m_*)$  and sum over  $m_*$  to obtain

$$\mathbb{P}_\phi^k \left\{ \frac{1}{n} \log D_\phi^{k,k+n} \xrightarrow{n \rightarrow \infty} I_\phi \right\} = 1. \quad (48)$$

For any sequence of number  $\{b_n\}_{n \in \mathbb{N}}$ ,  $\frac{1}{n} b_n \xrightarrow{n \rightarrow \infty} b$  implies that  $2 \frac{1}{N} \max_{1 \leq n \leq N} b_n \xrightarrow{N \rightarrow \infty} b$ . Thus, it follows from (48) that

$$\mathbb{P}_\phi^k \left\{ \frac{1}{N} \max_{1 \leq n \leq N} \log D_\phi^{k,k+n} \xrightarrow{N \rightarrow \infty} I_\phi \right\} = 1. \quad (49)$$

Since convergence a.s. implies convergence in probability, this implies (18).

---

<sup>2</sup>Here is the proof. Fix  $\varepsilon \in (0, b/2)$  and pick  $n_0$  so that for  $n \geq n_0$ ,  $|\frac{1}{n} b_n - b| \leq \varepsilon$ . Let  $B_p^q := \frac{1}{N} \max_{p \leq n \leq q} b_n$ . We have  $B_1^N = \max\{B_1^{n_0-1}, B_{n_0}^N\}$ . We can pick  $N_0$  such that for all  $N \geq N_0$ ,  $B_1^{n_0-1} \leq \varepsilon$ . On the other hand,  $n(b - \varepsilon) \leq b_n \leq n(b + \varepsilon)$ , for  $n \in [n_0, N]$ . Taking the maximum of each side over this interval, we obtain  $N(b - \varepsilon) \leq N B_{n_0}^N \leq N(b + \varepsilon)$ . Since  $2\varepsilon < b$ , we have  $B_1^N = B_{n_0}^N$  and  $|B_{n_0}^N - b| \leq \varepsilon$  which implies the result.

(b) To prove (19), let us fix  $\varepsilon \in (0, \varepsilon_0)$  throughout. Changing  $n$  to  $n-k+1$  in the definition of  $T_\varepsilon^k$ , we obtain

$$\begin{aligned} T_\varepsilon^k &= -k+1 + \sup \left\{ n \geq k : \frac{1}{n-k+1} \log D_\phi^{k,n} < I_\phi - \varepsilon \right\} \\ &= -k+1 + \sup \left\{ n \geq k : \frac{1}{n} \log D_\phi^{k,n} < \frac{n-k+1}{n} (I_\phi - \varepsilon) \right\} \\ &\leq -k+1 + \underbrace{\sup \left\{ n \geq 1 : \frac{1}{n} \log D_\phi^{k,n} < I_\phi - \varepsilon \right\}}_{\tilde{T}_\varepsilon^k}. \end{aligned}$$

Thus, it is enough to verify (19) for  $\tilde{T}_\varepsilon^k$  in place of  $T_\varepsilon^k$ .

Let  $Y^{k,n} := \frac{1}{n} \log D_\phi^{k,n}$ . For  $m_* \in \mathbb{N}^d$ , let  $n_2 := n_2(k, m_*; \varepsilon)$  be the smallest integer  $n$  that satisfies  $\sqrt{n} \geq \frac{1}{\varepsilon} p(m_*, k)$ , that is

$$n_2(k, m_*; \varepsilon) := \lceil \frac{1}{\varepsilon^2} p^2(m_*, k) \rceil \leq \frac{1}{\varepsilon^2} p^2(m_*, k) + 1. \quad (50)$$

Let  $n_0$  be as in the previous part. By assumption, for all  $n \geq \max\{n_2, n_0\}$ , we have  $\mathbb{P}_{\lambda_*}^{m_*} \{|Y^{n,k} - I_\phi| > \varepsilon\} \leq Cn^a \exp(-n\varepsilon^2)$ . To simplify notation, we will assume  $n_0 = 1$  without loss of generality. We have

$$\begin{aligned} \mathbb{E}_{\lambda_*}^{m_*} [\tilde{T}_\varepsilon^k] &= \sum_{\ell=1}^{\infty} \mathbb{P}_{\lambda_*}^{m_*} (\tilde{T}_\varepsilon^k > \ell) = \sum_{\ell \geq 1} \mathbb{P}_{\lambda_*}^{m_*} \left( \bigcup_{n > \ell} \{Y^{n,k} < I_\phi - \varepsilon\} \right) \\ &\leq \sum_{\ell \geq 1} \sum_{n > \ell} \mathbb{P}_{\lambda_*}^{m_*} (Y^{n,k} < I_\phi - \varepsilon) \\ &= \sum_{n=1}^{\infty} (n-1) \mathbb{P}_{\lambda_*}^{m_*} (Y^{n,k} < I_\phi - \varepsilon) \\ &\leq \sum_{n=1}^{n_2-1} (n-1) + \sum_{n \geq n_2} Cn^a e^{-n\varepsilon^2} \\ &\leq \frac{(n_2-1)^2}{2} + \sum_{n=1}^{\infty} Cn^a e^{-n\varepsilon^2}. \end{aligned}$$

The second term on the RHS does not depend on  $m_*$  or  $k$ , and we can denote it as  $C_1(\varepsilon)$ . Using the bound (50) on  $n_2$ , we have

$$\mathbb{E}_\phi^k [\tilde{T}_\varepsilon^k] \leq \frac{1}{2\varepsilon^4} \sum_{m_* \in \mathbb{N}} \pi_\phi^k(m_*) p^4(m_*, k) + C_1(\varepsilon).$$

Since by assumption, both  $\pi_\phi^k(\cdot)$  and  $\mathbb{P}(\phi = \cdot)$  have finite polynomial moments, it follows that (19) holds for  $\tilde{T}_\varepsilon^k$ .

## References

- [1] X. Nguyen, A. A. Amini, and R. Rajagopal, “Message-passing sequential detection of multiple change points in networks,” in *ISIT*, 2012.

- [2] T. L. Lai, "Sequential analysis: Some classical problems and new challenges (with discussion)," *Statist. Sinica*, vol. 11, pp. 303–408, 2001.
- [3] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Info. Theory*, vol. 39, no. 2, pp. 433–442, 1993.
- [4] A. M. Hussain, "Multisensor distributed sequential detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 698–708, 1994.
- [5] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2072–2089, 2008.
- [6] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On optimal quantization rules in some problems in sequential decentralized detection," *IEEE Transactions on Information Theory*, vol. 54(7), pp. 3285–3295, 2008.
- [7] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 534–548, 2011.
- [8] R. Rajagopal, X. Nguyen, S. Ergen, and P. Varaiya, "Distributed online simultaneous fault detection for multiple sensors," in *Proc. of 7th Int'l Conf. on Info. Proc. in Sensor Networks (IPSN)*, April 2008.
- [9] M. I. Jordan, "Graphical models," *Statistical Science*, vol. 19, pp. 140–155, 2004.
- [10] V. V. Veeravalli and T. Banerjee, "Quickest change detection," *E-Reference Signal Processing*, To appear.
- [11] S. Dayanik, C. Gouling, and H. V. Poor, "Bayesian sequential change diagnosis," *Mathematics of Operations Research*, vol. 33, no. 2, pp. 475–496, 2008.
- [12] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," in *Joint Statistical Meeting*, 2011.
- [13] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a markov process across a sensor array," *IEEE Transactions on Information Theory*, vol. 56(4), pp. 1961–1981, 2010.
- [14] M. Cetin, L. Chen, J. W. Fisher III, A. Ihler, R. Moses, M. Wainwright, and A. Willsky, "Distributed fusion in sensor networks: A graphical models perspective," *IEEE Signal Processing Magazine*, vol. July, pp. 42–55, 2006.
- [15] O. P. Kreidl and A. Willsky, "Inference with minimum communication: a decision-theoretic variational approach," in *NIPS*, 2007.
- [16] R. Rajagopal, X. Nguyen, S. Ergen, and P. Varaiya, "Simultaneous sequential detection of multiple interacting faults," <http://arxiv.org/abs/1012.1258>, 2010.

- [17] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [18] A. G. Tartakovsky and V. V. Veeravalli, “General asymptotic bayesian theory of quickest change detection,” *Theory Probab. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.
- [19] A. G. Tartakovsky and M. Pollak, “Nearly minimax changepoint detection procedures,” in *ISIT*, 2011.